Brain Magnetic Resonance Image Analysis: Segmentation and Registration

Liping Wang

Department of Computer Science
Aberystwyth University

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

February 2017

Declaration and Statement

DECLARATION

This thesis has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date .....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where **correction services**[1] have been used, the extent and nature of the correction is clearly marked in a footnote(s).
Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ..... (candidate)

Date .....

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date .....

[1] This refers to the extent to which the text has been corrected by others.

Acknowledgements

I am grateful for the opportunity and financial support from China Scholarship Council and Aberystwyth University, which makes my PhD study possible. First of all, I would like to express my great gratitude to my supervisor Prof. Reyer Zwiggelaar, who gave me enormous help from the very beginning of PhD application all the way to conducting research and writing up my thesis. He has always been patient to discuss with me about my ideas, listen to my doubts and provide insightful suggestions. Doing research is very challenging for me and I often felt frustrated in the past three years. His encouragement brought me through the darkness and desperation when I met difficulties or felt unconfident. I feel so blessed to have him as my PhD supervisor.

My thanks also go to my second supervisor Dr. Frédéric Labrosse, for his helpful comments on my papers and thesis. I would like to acknowledge the former and present members in Vision, Graphics and Visualisation Group, for their inspiring talks and discussions in research. Special thanks to Zhili Chen and Ziming Zeng, who helped and encouraged me a lot in the early stage of my PhD study.

I would like to express my appreciation to my dear friends Diana and David White House, Jean and Arthur Murray, for their love, care and prayers over all these years. Their enthusiasm to life and kindness to people have always been inspiring me. I learnt a lot from them about the Bible and British culture. Many thanks to my former and present housemates Leikai Ge, Anh, Juan Cao and Tianhua Chen, my house owners Paddy and Sue O'Brien and other Chinese friends. Their friendships have made my study at Aberystwyth a pleasant and memorable experience.

I would like to thank my best friends in China, for their continuous support throughout my PhD. The same goes to my boyfriend Ran Song, for sharing life together all these years. Finally, I would like to express my deepest gratitude to my family: my grandparents, parents and brothers. Thanks for their unconditional love, because of which I want to be better and stronger.

Abstract

Automatic and robust 3D brain magnetic resonance (MR) image analysis can assist in disease diagnosis, surgical planning and patient follow-up. As the most crucial techniques for brain image analysis, brain MR image segmentation benefits the pathology detection, volumetric morphometry, surface reconstruction and 3D visualisation; brain image registration helps with multi-modality image fusion, longitudinal analysis and population modelling. A variety of brain image segmentation and registration methods have been proposed. However, there is still room for improvement in terms of algorithm automation, accuracy and efficiency and the characteristics of existing methods for specific applications still can be more thoroughly investigated. Our work aims to obtain a good understanding in the underlying problems of brain image analysis, including brain segmentation, brain image registration and primary tissue classification, and propose possible improvements for the most popular and/or the most promising techniques. For brain segmentation, an improved Brain Extraction Tool (BET) method is proposed, which overcomes the weaknesses of the original method by enhancing the vertex displacement and embedding an independent brain surface reconstruction step during the iterative process of surface evolution. This strategy effectively deals with the surface self-intersection problem and results in faster algorithm convergence and better brain segmentation. For brain image registration, we propose a salient edge guided demons method, which uses salient edges detected in 3D scale-space rather than the whole image grid to drive the registration process. This method obtains statistically equal registration performance compared with the demons method using the whole image as demon points, while the execution time is dramatically reduced. For brain tissue classification, we compare the approaches using three main image based features: intensity, local prior and multi-atlas prior. The modelling of these features is described and the effectiveness of these features in brain tissue classification is investigated. This study provides a general guide on what image based features can be used for effective brain tissue classification.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Motivation

Brain disorders such as stroke and Alzheimer's disease are the leading causes of human deaths in the United States. Millions of Americans are diagnosed with Alzheimer's. And the cases are expected to triple by the year 2050 according to the report of the American Alzheimer's association in 2011 [22]. Since many brain disorders are chronic and incurable, they contribute significantly to the global disease burden in terms of patient suffering and economic cost. Furthermore, scientists have found that brain diseases are currently affecting people at earlier ages than ever before [224].

Three dimensional (3D) brain images generated by modern imaging technologies provide a non-invasive view of the internal brain structures. They assist with the disease diagnosis and surgical planning in clinic. Among all the imaging modalities, Magnetic Resonance Imaging (MRI) has drawn great attention due to its excellent tissue contrast and non-ionising radiation. However, it takes enormous human effort in the manual delineation and analysis of Magnetic Resonance (MR) images. The large amount of data makes evaluation a tedious task and prone to errors because of inter- and intra-variability.

Therefore, developing a computer-aided diagnosis system becomes a logical choice to improve the situation. Image segmentation and registration are the most crucial techniques in brain MR image analysis. Automated and accurate brain image segmentation benefits the pathology detection and volumetric morphometry in the diagnosis of some brain diseases such as multiple sclerosis (MS), Alzheimer's disease and schizophrenia. Image registration enables the radiologists to follow the evolution of pathologies in longitudinal sequences. It also helps with the patient's image analysis by fusion of prior knowledge or multi-modality images. Numerous brain image segmentation

and registration approaches have been proposed in the past decades [137, 151, 168]. However, there is still lots of room to improve the existing methods with respect to automation, accuracy, algorithm efficiency, etc. [137, 151, 290] and the characteristics of existing methods for specific applications still can be more thoroughly investigated.

1.2 Research Aims and Objectives

Our research aims to investigate the techniques for 3D brain MR image analysis, more specifically, for brain segmentation, brain image registration and tissue classification. We try to obtain a good understanding in the underlying problems, the existing approaches, their advantages and disadvantages. Then for the most popular and/or the most promising methods, we look into their weaknesses and propose possible improvements.

The objectives of our research are listed as follows:

- Conducting a thorough literature review on image segmentation, registration techniques and their evaluation methods, both for general images and for 3D brain MR images specifically.

- Improving the shortcomings of the original Brain Extraction Tool (BET) [263] for brain segmentation. Brain segmentation is a prerequisite of subsequent image analysis. Among other approaches, BET is the most commonly used and produces acceptable segmentation results. However, it requires a large number of iterations for the deforming brain surface to converge; it does not deal with the surface self-intersection effectively; and its segmentation accuracy still has room for improvement. These issues are addressed in our work.

- Developing salient edge guided demons for brain image registration. Demons is widely used for image registration because it provides a good compromise between the registration accuracy and the execution speed. The classic demons method uses the whole image grid as demon points to drive the registration process. In our work, we investigate the effect of using edges, especially salient edges, as demon points. Salient edges are extracted by extending Lindeberg's scale-space theory [170] into 3D.

- Comparing the approaches using different image information priors for brain tissue classification. Image intensity, local prior and multi-atlas prior are three

commonly used features for brain tissue classification. We elaborate the modelling of these priors and investigate their effectiveness in brain tissue classification. We also compare the performance of various multi-atlas label fusion methods.

1.3 Achievements

The achievements of our research presented in this thesis can be described as:

- An extensive literature review on generic image segmentation and registration methods and their evaluation measures was conducted. A more specific review on brain segmentation, brain image registration and tissue classification was also provided.
- An improved BET method was proposed for brain segmentation, which overcomes the disadvantages of the original BET method and results in more accurate brain segmentation and faster execution.
- A salient edge guided demons method was developed for brain image registration. Instead of using the whole brain image, the salient edges were extracted from the images in the scale space and used as demon points to drive the brain image registration. This method dramatically reduces the computation time without significantly deteriorating the registration accuracy.
- A comparative study was conducted on the three main image information priors used for brain tissue classification: the image intensity, the local and multi-atlas priors. This study provides a general guide on which kind of image information priors could be used for effective brain tissue classification.

1.4 Thesis Outline

The remaining chapters of this thesis are organised as follows:

- Chapter 2 provides a clinical background of 3D brain MR image analysis. A wide range of brain imaging technologies are introduced and compared in terms of physical principle and medical applications. The MRI technique is described in more detail. Subsequently the clinical applications of brain medical image analysis, including image segmentation, registration and classification, are

presented. Then existing challenges in brain MR image analysis are discussed and finally, the databases used for our experiments are described.

- Chapter 3 presents a technical literature review on generic image segmentation and registration approaches, their evaluation measures and the statistical analysis methods to compare the performance of different approaches.

- In Chapter 4, an improved BET method is proposed, which overcomes the disadvantages of the original method. The brain segmentation performance is compared with three other popular methods qualitatively and quantitatively on three databases with respect to four metrics.

- In Chapter 5, a salient edge guided 3D brain MR image registration method using demons is presented. Edges are extracted by extending Lindeberg's scale-space theory into 3D. The effect of using salient edges rather than other points to drive the demons registration process is investigated.

- In Chapter 6, a comparative study of brain tissue classification using different image information priors is conducted. Three main image information priors including image intensity, local prior and multi-atlas prior are modelled. The impact of the key parameters of each model are analysed. The effectiveness of the three priors in brain tissue classification are compared on both simulated and real databases.

- In Chapter 7, the work presented in the thesis is summarised. The contributions of the work are listed and future work is also discussed.

Chapter 2

Clinical Background

This chapter provides the background of brain MR image analysis from a clinical point of view. A wide range of brain imaging technologies are introduced. Subsequently the clinical applications of brain MR image analysis, including image segmentation, registration and classification, are described. Existing challenges in this area are presented and finally, the databases used for the experiments in the thesis are discussed.

2.1 Brain Imaging

In this section, we give a brief introduction to commonly used brain imaging modalities. Their physical principles, applications, advantages and limitations are presented and compared. The MR imaging mechanism is described in more detail since this is the imaging modality used in this thesis.

2.1.1 Brain Imaging Modalities: An Overview

Comparisons of various brain imaging modalities in terms of imaging principle, advantages and disadvantages are listed in Table 2.1. The image examples generated from these modalities are illustrated in Fig. 2.1.

Ultrasound Imaging (US) uses high-frequency sound waves to generate images of the inside of the body. The machine transmits sound waves into the body and converts the returning sound echoes into an image. The ultrasound exam is most commonly performed on infants to monitor brain conditions with prematurity [174]. It is also used to assess brain tumour masses in adults and the risk of stroke in both adults and children. Ultrasound scanning is noninvasive, unpainful, easy to use and extremely safe. On the other hand, ultrasound examinations are very sensitive to motion and

the exam results could be unstable due to an open wound or changes in blood flow pattern [76].

Computed Tomography (CT) uses a series of X-rays penetrating the brain from different directions [108]. A computer produces detailed cross-sectional images by estimating the amount of X-ray beams absorbed in small volumes of the brain. CT scanning provides the doctors 3D views of the bones, soft tissues and blood vessels [115]. It is painless, noninvasive and cost-effective. Compared with Magnetic Resonance Imaging, CT examination is faster and less sensitive to patient movement. However, the soft tissues in CT scans are less visible. Besides, excessive exposure to radiation has a slight chance to cause cancer. CT scanning is not recommended for children and pregnant women. It also has limitations in use for breastfeeding mothers [206].

MRI uses a powerful magnetic field and radio waves to generate 3D images of the brain structures. It does not use ionizing radiation. Owing to the better soft tissue contrast in MR images than other modalities [96], MRI possesses great value in early diagnosis of many brain diseases such as tumours and stroke. Nevertheless, the exam is more expensive and takes longer than other imaging techniques. The patients have to remain still and follow breath-holding instructions in order to produce high-quality images. As a result, MRI is not recommended for patients with acute injuries or claustrophobia. Moreover, because of the magnetic field used, patients cannot wear any metallic objects like pacemakers [118]. Pregnant women are not advised to take MRI exams during the first trimester [202].

Functional Magnetic Resonance Imaging (fMRI) is a functional neuroimaging technology which uses MRI to measure brain activities by detecting changes associated with blood flow [188]. It works based on the fact that blood flow changes with activation of brain neurons. fMRI shares the strengths and weaknesses of MRI and its clinical use lags behind research use [237]. In addition, it can be used to map the brain regions with functions associated with perception, thought and action so that brain regions which control important functions can be avoided in surgical planning. However, it is difficult to create fMRI images for patients with brain pathologies because tumours and lesions might cause blood changes unrelated to neural activities [237].

Positron Emission Tomography (PET) is a functional imaging technology which measures the emissions from radioactively labelled biologically active molecules that have been introduced into the bloodstream. The emission data is then collected to generate the 3D image which reveals the distribution of the molecules throughout the brain [205]. It provides information about brain activities in various regions

that relate to different rates of molecular uptake. It is valuable in diagnosing brain diseases [89], exploring cancer metastasis [261] and evaluating the effectiveness of treatments [197]. In particular it contributes significantly in early discovery of certain dementias (e.g. Alzheimer's and Pick's disease) where the early pathologies are not easy to be distinguished from normal tissues. PET scanning is noninvasive but involves exposure to ionizing radiation. Its main disadvantage is the high operation cost [59].

Single Photon Emission Computed Tomography (SPECT) is a functional nuclear imaging technique using gamma rays. It produces 3D information about brain function by measuring cerebral blood flow [7]. The image can be used to study brain diseases such as dementia, stroke, trauma, multiple sclerosis, etc. SPECT is less sensitive to patient movement compared with other modalities, which makes it well suited for epilepsy imaging. The application of SPECT is limited by its poor resolution [325]. The scanning is not recommended in situations of pregnancy and breastfeeding [269].

Electroencephalogram (EEG) is an electrophysiological monitoring test that measures and records the electrical activity of the brain. The electrodes, which are attached to the scalp with wires, analyse and send the electrical pulses from the brain to a computer. The computer records the electrical activity in the brain as dynamic waves. The changes in the normal pattern of brain electrical activity are detected to diagnose brain disorders such as epilepsy, brain tumour, memory disorders, stroke, dementia, etc. [98]. The EEG test is painless and safe. Its most important limitation is low spatial resolution [154].

Magnetoencephalography (MEG) is a neurophysiological technique that measures the magnetic fields generated by neuronal activity of the brain. The brain activity is localised by analysing the spatial distributions of the magnetic fields using very sensitive magnetometers. MEG has excellent spatial and temporal resolutions. It is completely non-invasive. Clinically, MEG is used to detect and localise pathological activity in patients with epilepsy and study cognitive processes in fetuses and newborns [257]. The main drawback of MEG is that the signals of interest are extremely small and thus specialised shielding is required to eliminate the magnetic interference in a typical environment.

Table 2.1 Comparisons of brain imaging modalities

Category	Modality	Imaging principle	Advantages	Disadvantages
Anatomical	US	high-frequency sound waves	noninvasive, painless, safe and easy to use	motion sensitive, unstable examining results
	CT	multiple X-rays	noninvasive, painless, cost-effective, fast, less sensitive to motion	exposure to X-Rays, not recommended for children, pregnant women or breastfeeding mothers
	MRI	powerful magnetic field and radio waves	noninvasive, better spatial resolution, excellent tissue contrast	expensive, long examining time, motion sensitive, not recommended for patients with claustrophobia or metallic objects
Functional	fMRI	blood oxygenation level dependent effect	noninvasive, enables the detection of brain abnormalities and assessment of normal functional anatomy	not recommended for patients with claustrophobia, pathologies or metallic objects
	PET	emissions from radioactively labelled chemicals	noninvasive, early discovery of certain dementias	exposure to ionizing radiation, high cost
	SPECT	gamma rays	less sensitive to patient motion	poor resolution, contraindicated in situations of pregnancy and breastfeeding
	EEG	electrical pulses in the brain	painless, safe	low spatial resolution
	MEG	magnetic fields generated by brain activity	completely noninvasive, excellent spatial and temporal resolutions	specialised shielding is required

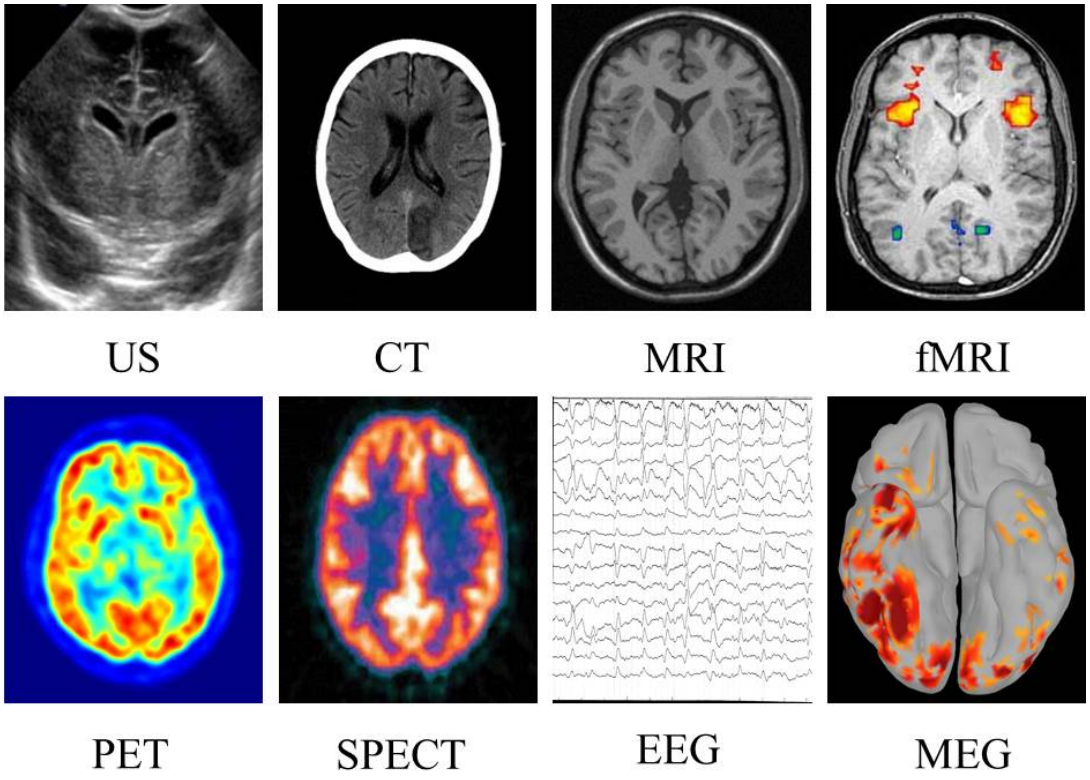


Fig. 2.1 Brain image examples generated by different imaging technologies [289, 78, 46, 111, 218, 270, 97, 194]. These images are used with permission.

2.1.2 Brain Magnetic Resonance Imaging

Physical principles

The first MRI exam was performed on a human being in 1977 by Dr. Raymond Damadian and his colleagues [118]. It works owing to the abundant hydrogen atoms contained in water, which makes up most of the human body. An MRI scanner has several important components: the bore, magnet, radio frequency coil, three gradient coils and a receiver coil, as shown in Fig. 2.2. Under natural conditions, the hydrogen atoms are randomly spinning in our body. During the scan, the magnet creates a magnetic field of 0.5-Tesla to 3.0-Tesla (although higher field strength research falicilities are available). When the patient enters the bore, the hydrogen atoms line up in the direction of the field. The majority of the hydrogen protons pointing to opposite directions cancel each other out to remain in equilibrium state. Then the radio frequency coil transmits pulse toward the area of the body being scanned. This pulse knocks the unmatched protons out of alignment and makes them spin at the Larmour frequency [331]. In the meanwhile, three gradient magnets are turned on and off

frequently to change the main magnetic field at a local level, which allows the scanner to pick the particular area of interest and generate cross-sectional images. When the radio frequency pulse ceases, the unmatched protons realign and release radio signals, which are collected by the receiver coil and sent to a computer to produce a detailed 3D image of the internal structures or tissues of the human body [23].

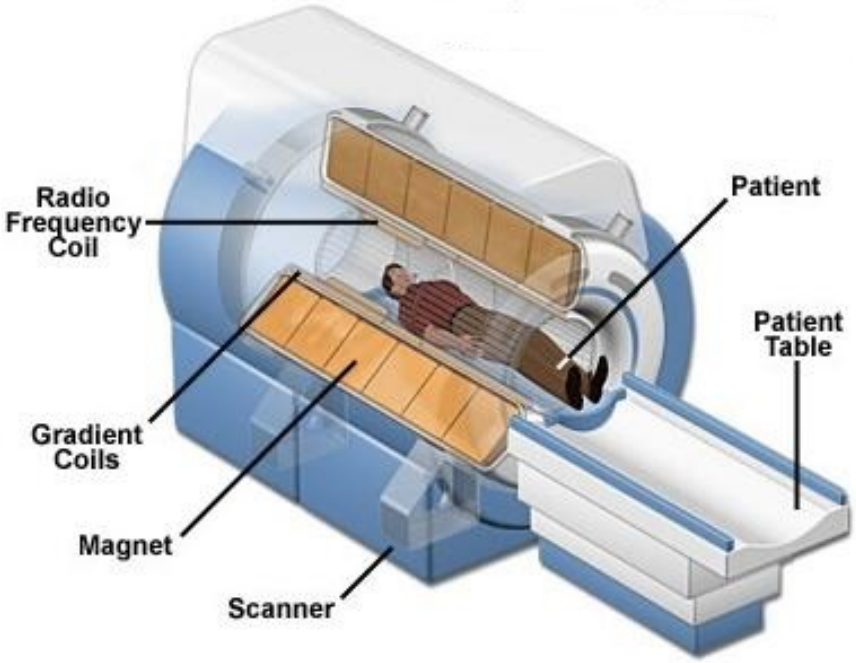


Fig. 2.2 A cutaway of MRI scanner [203].

Brain MRI sequences

A variety of MRI sequences with different tissue contrast can be produced by adjusting the parameters of MR scanners such as repetition time (TR) and echo time (TE). TR is the time between the application of successive radio frequency pulses. It determines the extent to which the protons in tissues to relax back into alignment with the main magnetic field [246]. TE is the time between the application of the radio frequency pulse and the peak of the signal induced in the coil [39]. It controls the amount of the signal decay that occurs before the signal is received. Both TR and TE are measured in milliseconds.

At short TR, the differences in relaxation time among various tissues can be detected and a T1-weighted (T1-w) image is generated. If the TR is too long, all the tissues recover their longitudinal magnetisation completely. Then the relaxation time differences are not demonstrated in the image [39]. In brain T1-w images (see

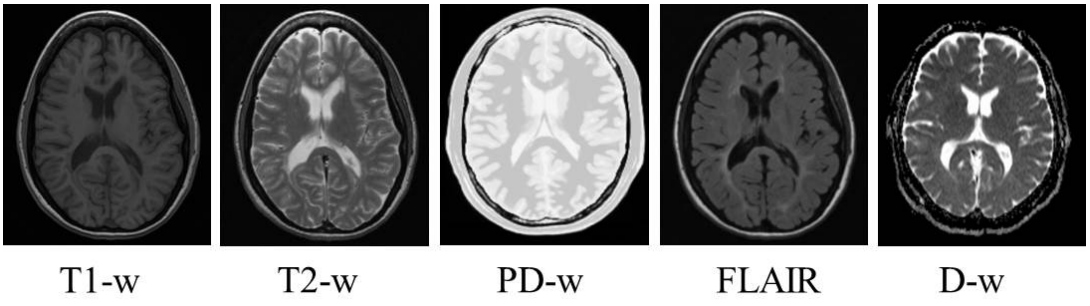


Fig. 2.3 Example of brain MR images of various sequences. T1-w, T2-w and FLAIR images are from MICCAI 2008 MS lesion segmentation challenge [253]; PD-w image is from [46]; D-w image with ADC mapping is from [88].

Fig. 2.3), the white matter (WM) has high signal intensity and appears bright; the grey matter (GM) has intermediate signal intensity and appears grey; the cerebral spinal fluid (CSF) has low signal intensity and appears dark.

To generate T2-weighted (T2-w) images, the TEs must be long enough to allow the tissue signals to decay. If TEs are too short, the differences in the signal decay among different tissues cannot be detected. In brain T2-w images (see Fig. 2.3), the signal intensities of WM, GM and CSF show in the opposite way to those in T1-w images: WM is dark; GM still has intermediate intensity; CSF is bright.

When the TR is long and TE is short, the tissue contrast resulting from the differences in both magnetisation recovery and signal decay is diminished and thus proton density weighting can dominate [39]. The tissue contrast in proton density weighted (PD-w) images is formed owing to the differences in proton density among tissue types. From an example of brain PD-w image in Fig. 2.3, it is seen that CSF is bright and GM is brighter than WM.

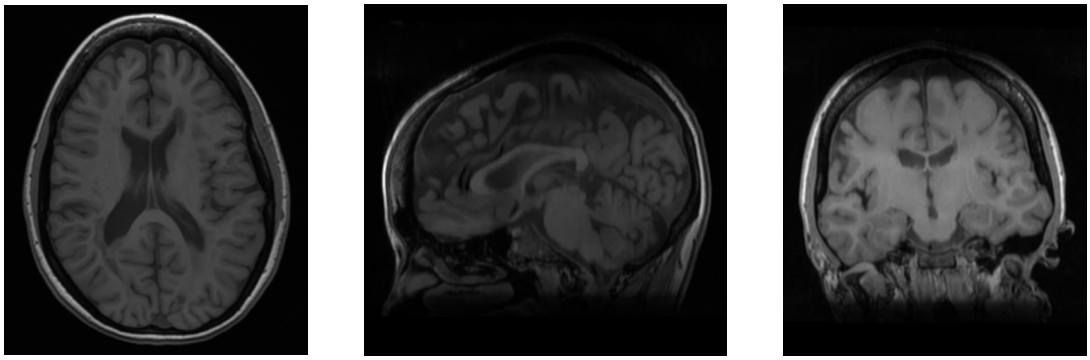
Fluid attenuation inversion recovery (FLAIR) is a special inversion recovery sequence which applies 180 degree and 90 degree radio frequency pulses successively. It is often used to suppress the effects of fluid from the resulting images by adjusting the inversion time (the time between the application of two radio frequency pulses) to the zero crossing point of fluid. Brain FLAIR imaging is useful in diagnosing diseases of the central nervous system such as cerebral infarction and MS lesion [27]. An example of brain FLAIR images is shown in Fig. 2.3 in which CSF is suppressed and MS lesions are highlighted.

Diffusion-weighted (D-w) imaging enables us to distinguish the tissue types according to the diffusion speed of protons. In the human brain, the diffusion of CSF is unrestricted; GM and WM possess intermediate diffusion speed. One of the major applications of brain D-w imaging is the diagnosis of recent stroke. It is usually applied

in conjunction with apparent diffusion coefficient (ADC) mapping techniques [39]. A brain D-w image with ADC mapping is shown in Fig. 2.3.

Brain MR image views

Brain MR images are commonly viewed in three planes: axial, sagittal and coronal. As shown in Fig. 2.4, axial slices are taken from the bottom to the top of the head; sagittal slices are taken from side to side; coronal slices are taken from the front to the back of the head.



(a) Axial (b) Saggital (c) Coronal

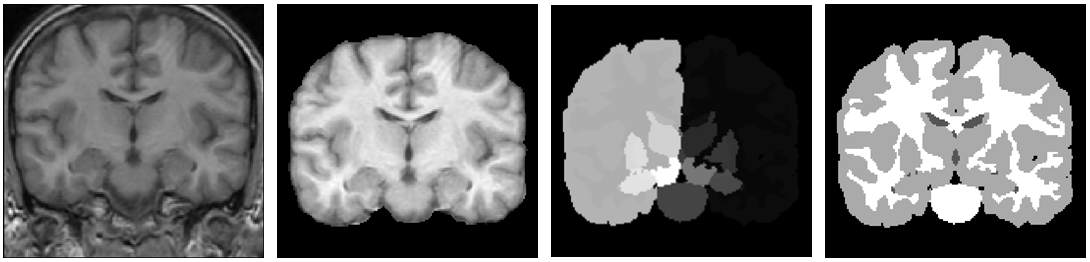
Fig. 2.4 Brain MR images viewed in three planes. Three brain slices are from images of MICCAI 2008 MS lesion segmentation challenge [253].

2.2 Applications of Brain Magnetic Resonance Image Analysis

In this section, we introduce the applications of brain MR image analysis with regard to segmentation, registration and classification.

2.2.1 Segmentation

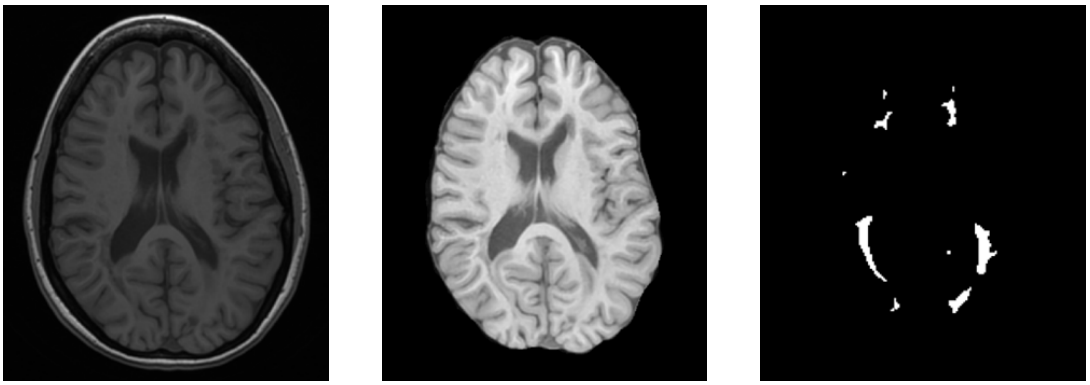
Brain segmentation is more commonly referred to as skull stripping. It aims to segment brain tissues including WM, GM and CSF from non-brain tissues such as the skull, scalp, eyeballs, muscle, etc., (Figs. 2.5 (b) and 2.6 (b)). Accurately removing redundant tissue could improve the robustness of subsequent processing steps such as bias field correction, registration between functional and MR images, segmentation of internal structures, tissues and pathologies. Moreover, some applications such as



(a) (b) (c) (d)

Fig. 2.5 Coronal slices of brain volume from IBSR [135]: the original image (a), the skull stripped image (b), the segmented structures (c) and tissues (d).

cortical surface reconstruction, cortical thickness measurement [291] and brain atrophy estimation [264] can also benefit from skull stripping.



(a) (b) (c)

Fig. 2.6 Axial slices of brain volume from MICCAI 2008 MS lesion segmentation challenge [187]: the original image (a), the image after skull stripping (b) and the segmented MS lesion (c).

Structure segmentation aims to segment the internal structures or tissues from brain images. For example, Fig. 2.5 (c) shows the segmented brain structures including left and right caudate, putamen, ventricle, hippocampus, etc.; Fig. 2.5 (d) shows the segmented brain tissues including WM, GM and CSF. It assists in some medical image analysis applications like image registration [128], lesion segmentation [143] and cortical surface extraction [149]. Brain structure segmentation is an indispensable prerequisite for 3D volume visualisation. It is used for voxel based morphometry to detect subtle changes in brain structures, which are associated with neurodegenerative and psychiatric diseases [193, 109]. For instance, early and disproportionate hippocampal atrophy is characterised in people diagnosed with Alzheimer's disease; focal atrophy of the temporal or frontal lobe is more likely to be linked with Pick's disease [20]. Such structural changes are considered as biomarkers to track the disease progress.

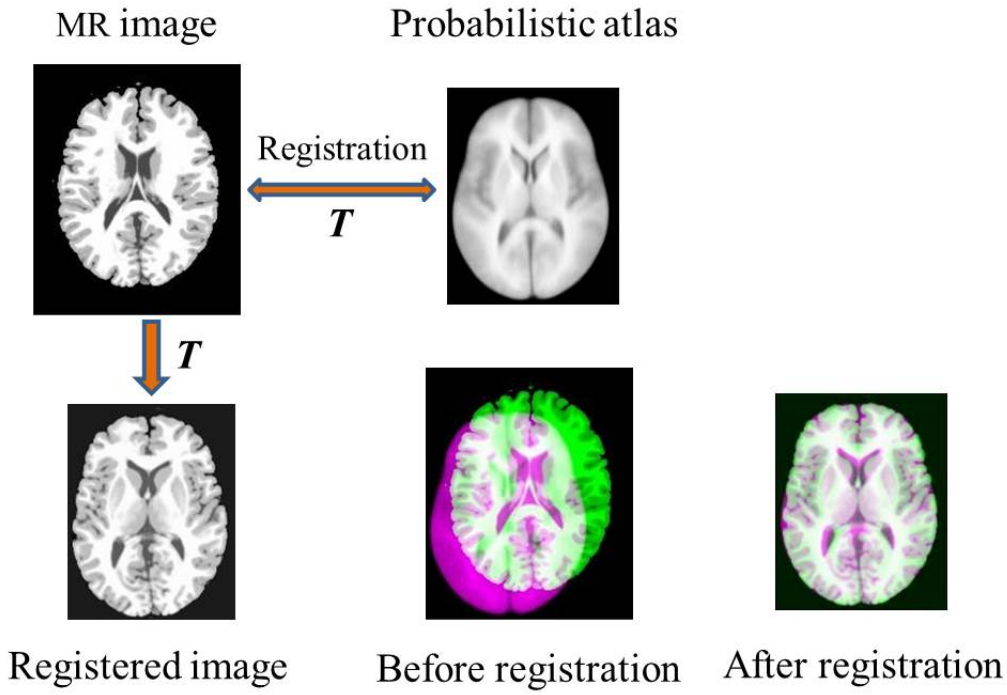


Fig. 2.7 Registration between an MR T1-w image and a probabilistic atlas. The composite RGB images show the atlas and MR image before and after registration overlaid in different colours. Gray colour shows the regions where two images have the same intensities; Magenta and green colours label the regions where the intensities are different. The MR image is a simulated image from [46] and the atlas is from the LONI Probabilistic Brain Atlas (LPBA40) [253].

A selective review is provided in [11] to demonstrate the findings of the relationship between brain structure and neurocognition in schizophrenia. Clinically, segmentation of these brain structures could help in disease diagnosis, surgical planning and patient follow-up. It is also used in studies of healthy subjects to investigate the impact of learning and practice activities on the changes of brain structures [193].

Pathological region segmentation tries to segment brain tumours or focal lesions (e.g. MS lesions as shown in Fig. 2.6 (c)), which cause dementia, epilepsy, paralysis and headache [91]. It benefits in 3D visualisation and quantitative analysis of brain pathological regions. Accurate segmentation provides surgical guidance in surgical planning, volumetric analysis, time series analysis and computer aided diagnosis [114].

2.2.2 Registration

Brain image registration aims to find a transform that brings one image into a voxel-to-voxel correspondence with another image. In medical image analysis, it is often

used for multi-modality fusion which efficiently integrates information obtained from various devices or protocols [123]. Registration between anatomical and functional images enables the pathologists to locate the region of interest more precisely and better understand the brain function in a particular structure. It can also be used for observing the evolution of pathological tissues or evaluating the treatment effects by conducting longitudinal analysis [296]. Image registration is crucial in applications where the brain anatomy deforms over time. It is used in correction of artefact due to patient motion and geometric distortion in MR images [235]. Furthermore, it is a necessary process for population modelling like probabilistic atlas construction [38] and propagating the prior knowledge embodied in the atlases onto other medical images [79]. For example, registering to a topological atlas benefits applications where the brain shape and structure connectivity plays an important role [32]; the tissue or structure probabilities in a probabilistic atlas facilitate the analysis of the target image; the labels of multiple atlases can be used to infer the segmentation of the target image by applying some label fusion strategies [173]. Fig. 2.7 illustrates the registration process between a normal MR T1-w image and a probabilistic atlas. The transformation T is calculated between the two images. The registered image is then obtained by applying T to the original T1-w image.

2.2.3 Classification

Brain image classification refers to classifying clinical images into different categories according to the type or degree of the diseases. Accurate and automatic brain image classification assists in computer aided diagnosis and reduces the human effort in making decisions. In the simplest cases, brain images are classified as either normal or abnormal [60]. Brain images are also classified into categories with various diseases such as glioma, meningioma, Alzheimer's disease, Pick's disease or Huntington's disease [334]. In addition, one type of brain disease like MS lesions can be further classified into several subtypes including enhancing lesions, T1 'black holes' and T2 hyperintense lesions [321, 200]. Fig. 2.8 shows brain MR T1-w images with three degrees of MS lesions.

2.3 Existing Challenges

Research on brain MR image analysis has been studied for decades and a variety of approaches have been proposed. However, there is no generic method that can solve the problems perfectly. The accuracy and robustness of the methods suffer from the

(a) (b) (c)

Fig. 2.8 Axial slices of brain volumes with mild (a), moderate (b) and severe MS lesions (c). MS lesion voxels are labelled with the brightest color. Images are from BrainWeb [46].

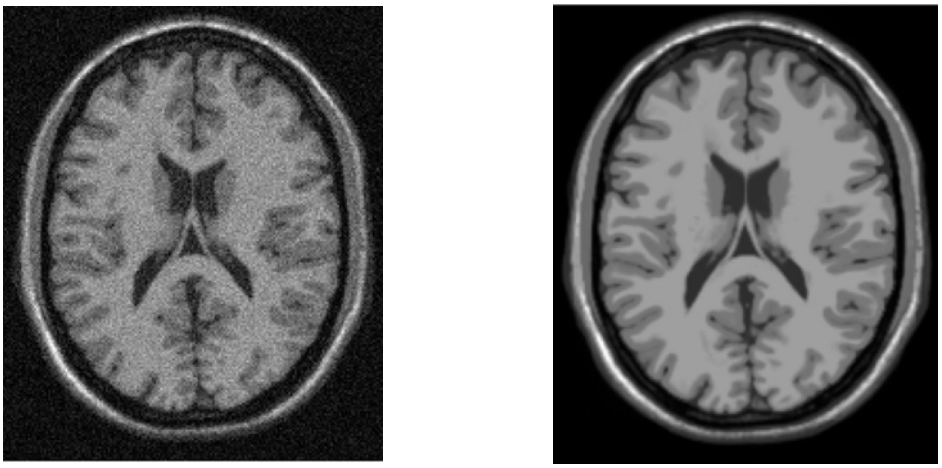
inherent artefacts that exist in MR images due to the scanner property and the imaging process. Generally these artefacts include noise, intensity inhomogeneity and partial volume effect. They are considered as the open problems in brain MR image analysis. These problems could be mitigated by improving the imaging techniques and applying some postprocessing steps to the images.

2.3.1 Noise

MR images are usually corrupted by random noise produced in the acquisition process. The noise is either Gaussian or Racian distributed with uniform or nonuniform variance across the image. Severe noise can affect the sharpness of the image or blur the boundaries of tissues. Numerous image denoising methods have been developed from traditional filtering methods [61, 100] to advanced anisotropic diffusion filters [217, 113]. In recent years, the non local means algorithms [50, 183] and wavelet based image denoising techniques [175, 241] have been shown to be successful in MR image denosing. Fig. 2.9 shows a comparison of images with and without noise.

2.3.2 Intensity Inhomogeneity

Intensity inhomogeneity, which is also referred to as intensity nonuniformity or bias field, is a common phenomenon in brain MR images. Due to the imaging acquisition system or patient movement, the intensity varies smoothly over the image domain (see Fig. 2.10). Owing to this phenomenon, the intensity of the same tissue shows variation at different locations in the image. It degrades the accuracy of the subsequent



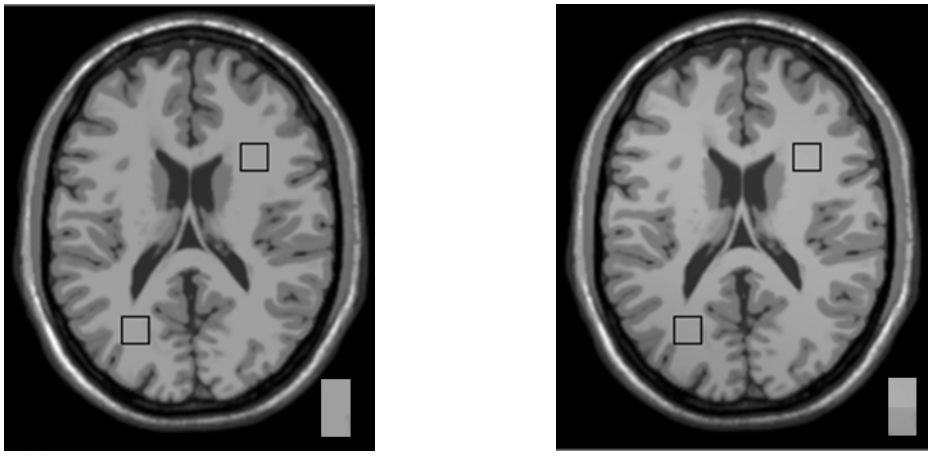
(a) (b)

Fig. 2.9 Simulated brain MR images with severe noise (a) and without noise (b). Images are from BrainWeb [46].

quantitative image analysis. In its simplest form, intensity inhomogeneity is modelled as multiplicative or additive to the image intensities. A number of methods have been proposed to correct intensity inhomogeneity in the last two decades. Generally, they can be divided into prospective and retrospective approaches [299]. The prospective approaches are further classified into phantom, multicoil and special sequences based methods. The retrospective methods include filtering, surface fitting, segmentation and histogram based algorithms.

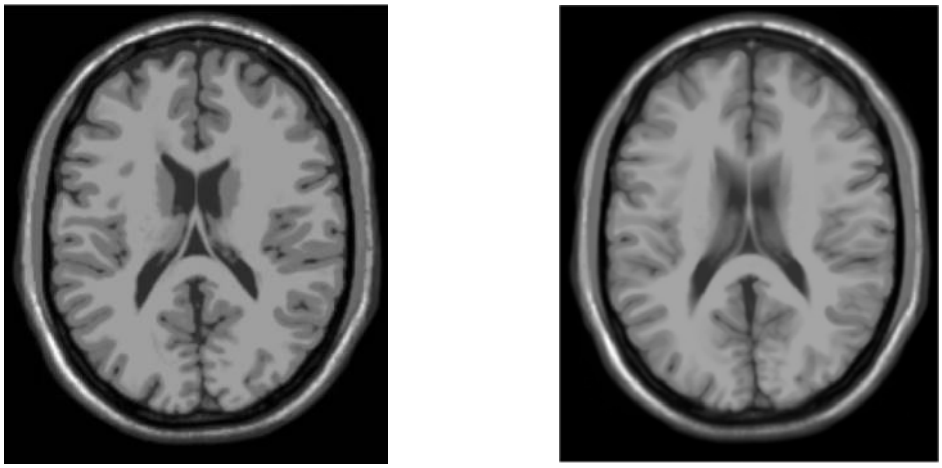
2.3.3 Partial Volume Effect

Partial volume effect indicates that multiple tissue types exist in a single voxel due to the finite spatial resolution of the scanner or the image reconstruction process. This phenomenon can also result from image sampling or respiratory motion [99]. It occurs at the interface of multiple tissue types. Fig. 2.11 shows simulated images generated at different resolutions. The tissue boundaries in the low-resolution image (Fig. 2.11(b)) are severely blurred because of partial volume effect. In brain MR image analysis, the quantitative measurement of tissues could be biased if this effect is not taken into account. The difference between the hard segmentation of GM and the ground truth shown in Fig. 2.12 is clear. In the literature [99, 266], partial volume effect is addressed by applying image enhancement, image-domain correction and projection-based correction techniques.



(a) (b)

Fig. 2.10 Simulated brain MR images without bias field (a) and with high level of bias field (b). The grey bar at the bottom right of each image illustrates the two patches extracted from the regions labelled by the rectangles in each image. The locations of the rectangular regions in the two images are the same. Original images are from BrainWeb [46].



(a) (b)

Fig. 2.11 Simulated brain MR images with slice thickness of 1*mm* (a) and 9*mm* (b). Images are from BrainWeb [46].

(a) (b)

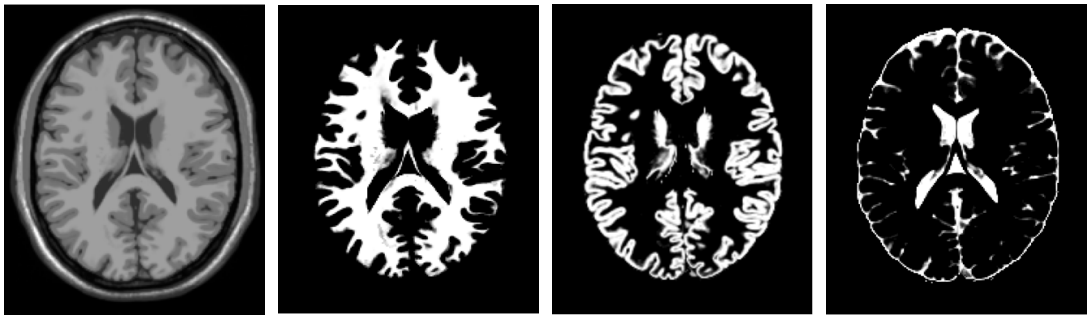
Fig. 2.12 Hard segmentation of GM (a) and the ground truth (b). Original images are from BrainWeb [46].

2.4 Databases

Both simulated and real brain MR image databases are used for the experiments in this thesis. All of them are publicly available. In this section, we will introduce each of them in detail.

2.4.1 BrainWeb

BrainWeb is a simulated brain MR image database [46]. The anatomical model was used to simulate images with different parameters. At present, BrainWeb provides images simulated using three pulse sequences (T1-w, T2-w and PD-w), 5 slice thicknesses (1mm, 3mm, 5mm, 7mm and 9mm), 6 noise levels (0%, 1%, 3%, 5%, 7% and 9%) and 3 levels of intensity inhomogeneity (0%, 20% and 40%). The size of the brain volume is $181 \times 217 \times 181$ and the image in-plane resolution is $1mm \times 1mm$. For the anatomical model, the fuzzy models of 9 tissue types (WM, GM, CSF, skull, fat, muscle, skin, glial matter and connective) are available. The voxel values in each fuzzy model give the proportion of the tissue in the voxels. A discrete anatomical model with hard tissue classification is also provided. These fuzzy models and the discrete model can be considered as ground truth in experiments. Because of the accessible ground truth and abundant parameters of the simulated images, BrainWeb is widely used to validate the image analysis methodologies (e.g. skull stripping, tissue classification) and investigate the robustness of algorithms to image artefacts. In addition to the normal brain database, BrainWeb also provides a MS lesion database, which can be used to evaluate MS lesion segmentation approaches. The limitations of



(a) (b) (c) (d)

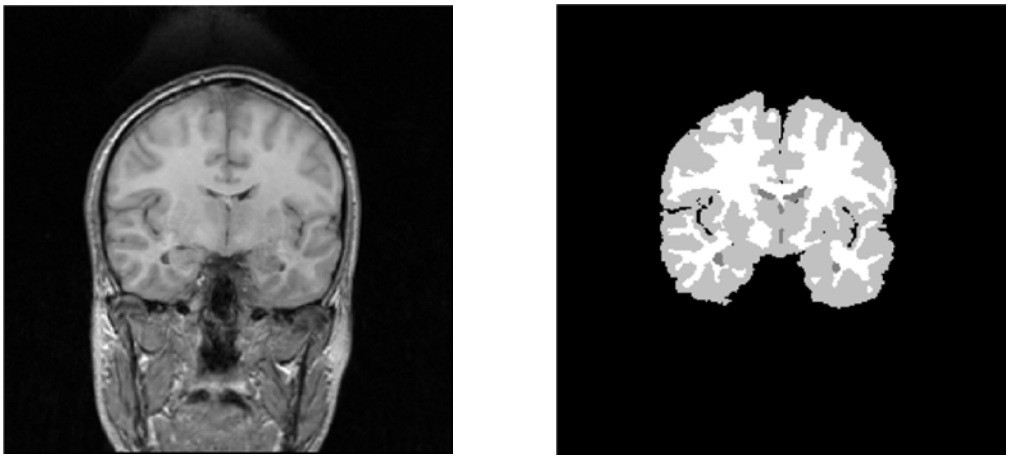
Fig. 2.13 Normal anatomical model (a) and fuzzy models of WM (b), GM (c) and CSF (d) in BrainWeb.

BrainWeb lie in the facts that the one phantom cannot take the anatomical variability into account; the simulated images are simpler than real images; other useful sequences such as FLAIR are not available and the MS lesion database is very limited compared to the real cases. In recent years, 20 normal anatomical models have been released on BrainWeb, which increases the variety of normal brains. Fig. 2.13 illustrates the normal anatomical model and fuzzy models of the primary tissues.

2.4.2 IBSR

The Internet Brain Segmentation Repository (IBSR) datasets and their manual segmentations were provided by the Center for Morphometric Analysis (CMA) at Massachusetts General Hospital [135]. The database, released in 1997, consists of 20 T1-w normal brain images with their manual segmentations. The size of each volume is typically $256 \times 256 \times 61$ with the resolution $1mm \times 1mm \times 3mm$. The images were positionally normalised and the segmentation was performed by trained experts using a semi-automated intensity contour mapping algorithm [146] and signal intensity histograms. Fig. 2.14 shows an example of original image and tissue classification in IBSR.

Application of the database described above is limited by its poor spatial resolution. During 2003 and 2004, a high-resolution database was published by the IBSR project. It consists of 18 T1-w normal brain images with their segmentations. The size of each image is $256 \times 256 \times 128$ with the resolution varying between $0.8mm \times 0.8mm \times 1.5mm$ and $1mm \times 1mm \times 1.5mm$. These images have been 'positionally normalised' into the Talairach orientation and bias field corrected by the CMA 'autoseg' routines. Each brain volume is skull stripped. Segmentations of 84 structures and three tissues are provided as shown in Fig. 2.5(c) and (d).



(a) (b)

Fig. 2.14 Coronal slices of a brain volume (a) and tissue classification (b) from IBSR. Three tissues (WM, GM, CSF) are labelled in different colours.

In the rest of this thesis, these two databases are referred to as IBSR20 and IBSR18, respectively. The IBSR databases are mostly used to validate segmentation and registration methods. Beyond that, these 3D volumes can also be used either as a set of topological atlases or to construct a probabilistic atlas after co-registering all the segmented images to a standard coordinate system and computing the frequency of each voxel belonging to a specific tissue or structure.

2.4.3 LPBA40

The LONI Probabilistic Atlas (LPBA40) provides maps of brain anatomical regions [253]. Forty whole-head MR images were taken from 40 healthy volunteers aged between 16 and 40 with equal number of males and females. The brain volumes were acquired on a GE 1.5T system with resolution of $0.86mm \times 0.86mm \times 1.5mm$ (38 subjects) or $0.78mm \times 0.78mm \times 1.5mm$ (2 subjects). The brain volumes were rigidly aligned to the MNI-305 average brain [103], which is considered as the delineation space with resolution $1mm \times 1mm \times 1mm$. Then the aligned images were bias field corrected using a non-parametric inhomogeneity normalisation algorithm [262] and skull stripped using the BET [263]. Following the defined protocols, 56 structures were labelled by 15 trained raters using the BrainSuite software package [251]. The delineated brain images were then aligned to the canonical atlas space by applying the delineation-to-native and native-to-atlas transforms calculated from the registration processes [317, 265, 19]. Finally, for each structure, 40 delineated volumes transformed to the atlas space were averaged to produce the probability density maps, which

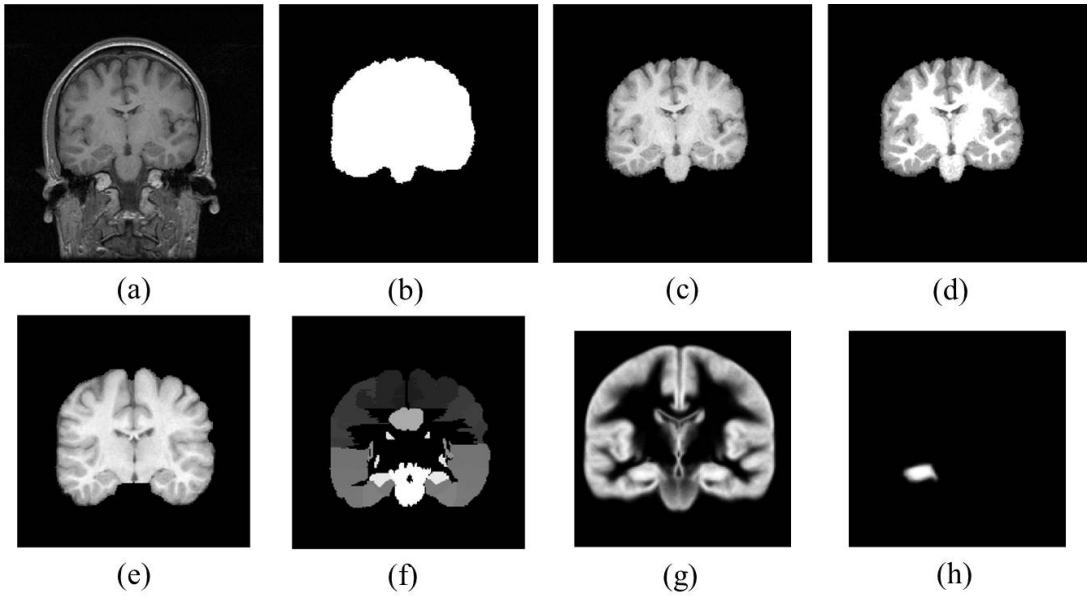


Fig. 2.15 Coronal brain slices of LPBA40 in the native space (a-d), delineation space (e, f) and atlas space (g, h). These images are the original brain image (a); the brain mask (b); the bias field corrected brain image (c); the tissue classification (d); the skull stripped brain image (e); the structure delineation (f); the probabilistic atlases of GM (g) and left hippocampus (h).

are referred to as probabilistic atlases. Fig. 2.15 illustrates the brain images of LPBA40 in the native space (with volume size $256 \times 256 \times 124$), delineation space (with volume size $181 \times 217 \times 181$) and atlas space (with volume size $149 \times 188 \times 148$). The abundant data LPBA40 provided facilitates brain image analysis in several applications. Tissue and structure probabilistic maps provide prior information in segmentation tasks. The 40 delineated brain volumes can be used to validate structure segmentation and registration methods. The delineated brain volumes can also serve as multiple atlases to provide prior knowledge in structure segmentation approaches or considered as a training dataset in machine learning based algorithms.

2.4.4 MICCAI2013

MICCAI2013 is a clinical brain image database used in the Grand Challenge on MR Brain Image Segmentation workshop at the international conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) in 2013 [195]. This challenge aims to provide an online evaluation framework for automatic algorithms for segmentation of WM, GM and CSF. Twenty 3T scans are available including 5 cases for training and 15 for testing. The images were acquired at the UMC Utrecht from patients aged over 50 with diabetes, varying degrees of atrophy and white

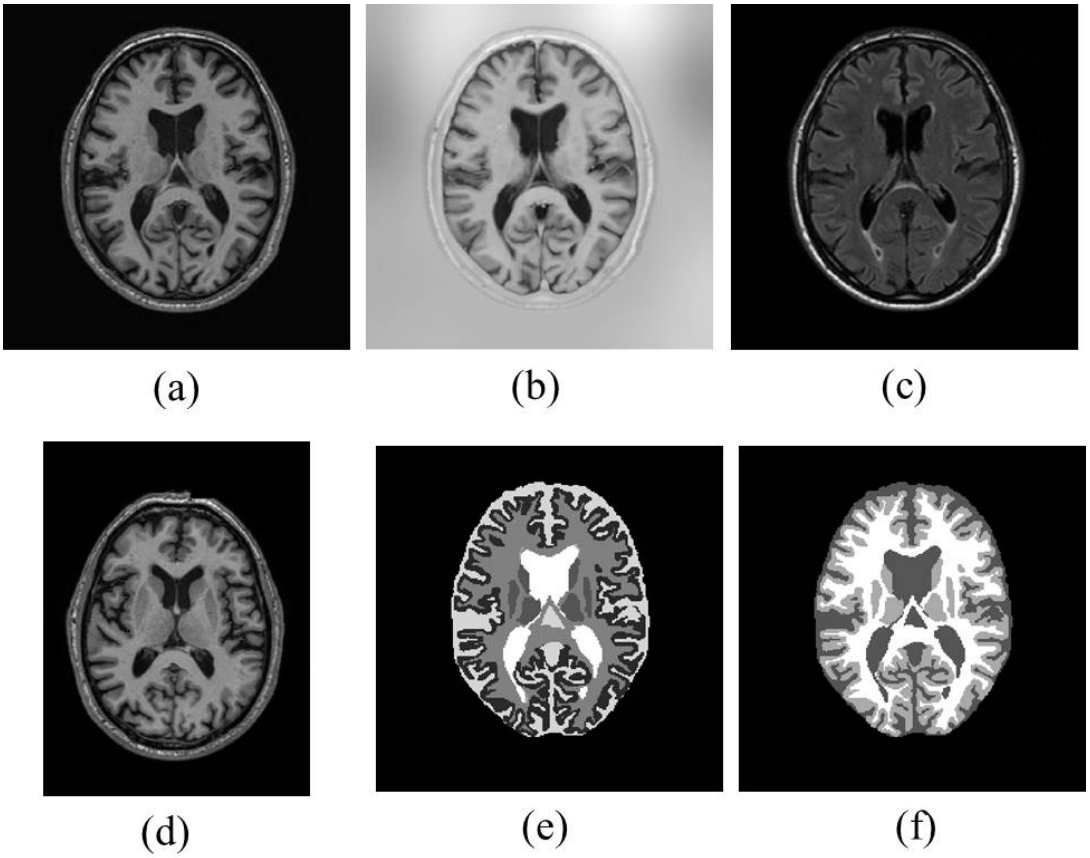


Fig. 2.16 Multi-sequence axial slices of brain volumes in training sets of MICCAI2013. Images are T1-w (a), T1-w IR (b) and FLAIR (c) slices; high-resolution T1-w image (d); manual segmentations of structures (e) and tissues (f), respectively.

matter lesions. Multi-sequence brain images are provided as shown in Fig. 2.16 (a-c) including T1-w, T1-w inversion recovery (IR) and FLAIR scans with volume size $240 \times 240 \times 48$ and resolution $0.958mm \times 0.958mm \times 3mm$. Besides, high-resolution T1-w scans are also available (see Fig. 2.16 (d)) with volume size $256 \times 256 \times 192$ and resolution $1mm \times 1mm \times 1mm$. All scans were bias corrected and the thick-slice ($3mm$ slice thickness) T1-w and T1-w IR images were aligned to the FLAIR images. For 5 training cases, manual segmentations of 8 brain structures and 3 main tissue types were completed (see Fig. 2.16 (e, f)) using techniques based on the contour segmentation objects tool from Mevislab [196]. This database can be used for validation of brain segmentation, structure and tissue segmentation methods on multi-modality clinical images.

2.5 Summary

This chapter presents the clinical background of brain MR image analysis. A variety of brain imaging techniques have been briefly described and compared in terms of the physical principles, advantages, disadvantages and applications. A more detailed introduction to brain MRI has been given. The applications of brain MR image analysis including image segmentation, registration and classification have been described. We also presented the existing challenges for brain MR image analysis. Finally, the simulated and real brain MR image databases used for our experiments in the thesis have been introduced.

Chapter 3

Technical Literature Review

In this chapter, a literature review on generic image segmentation and registration techniques, their evaluation algorithms and statistical analysis methods is conducted. More specific literature review on the brain segmentation, brain image registration and tissue classification techniques has been provided in Section 4.1.1, 5.1.1 and 6.1.1, respectively.

3.1 Image Segmentation Algorithms

A great number of image segmentation algorithms have been developed over many years in the literature. Most of them have been applied on medical images covering various applications. According to the principles of the methodologies, we classify them into six categories: thresholding, region growing, deformable models, machine learning techniques, graph theory based methods and atlas based approaches. For algorithms in each category, their principles, advantages, limitations and a few applications are elaborated.

3.1.1 Thresholding

Thresholding is a popular tool for segmenting the foreground of an image from the background. The threshold regarding the image feature (e.g. intensity, gradient magnitude) is determined manually or automatically. Then the voxels with image feature greater than the threshold are segmented as foreground and the others are background. Based on the strategy used to find the threshold, the algorithms can be further grouped into global, local and multithresholding techniques.

Global thresholding algorithms determine a global threshold for the entire image based on prior knowledge, histogram analysis or information theory. In the cases where the percentage of voxels belonging to the foreground can be estimated, a simple percentage (p-tile) method could be applied to define the threshold as the highest grey level, which maps at least $(100-p)\%$ of the voxels into the foreground in the thresholded image [245]. For the images with distinct foreground and background, the threshold can be chosen by histogram analysis. For binary segmentation, two separate peaks in the image histogram are considered as the objects and background. And the minimum grey level between the peaks is defined as the threshold which could be calculated by the mode method [245], Otsu's method [209] or the Isodata method [232]. Based on the grey value histogram of the image, thresholding can also be performed by applying information theory. The threshold is defined by either maximising the upper bound of the a posteriori entropy [225, 144] or minimising the interdependence between two separate peaks of the histogram [140]. A Bayesian thresholding approach was also proposed, which assumes the grey values of voxels belonging to the foreground and background follow Gaussian distributions with different means but equal variations [184]. The application of global thresholding is restricted by the lack of prior knowledge, extremely unequal peaks or broad and flat valleys in the image histogram, noise and intensity inhomogeneity.

Local thresholding compensates some shortcomings of global thresholding by taking the neighbouring voxels in a local region into account. Histogram transformation methods endeavour to generate a new histogram of the image with sharper peaks and deeper valleys by weighting the voxels in the image according to the local property. Then the global thresholding approaches based on histogram analysis can be easily applied on the transformed histogram [245]. The commonly used weighting schemes include various edge detectors (e.g. Canny, Sobel and Laplacian edge detectors) and quadtree methods [318]. In addition to the local weighting scheme, a spatially varying threshold can be determined by calculating the local mean and standard deviation [204] or estimating the value of current voxel using linear combination of the neighbouring voxels [184, 207]. Furthermore, methods based on second-order grey level statistics such as the co-occurrence matrix method [126] and the scatter plot method [150] also define the threshold using neighbouring voxels.

Multithresholding applies thresholding more than once in the segmentation approach. A hysteresis strategy was proposed in [56], which uses two thresholds to perform a two-stage thresholding. A recursive scheme is often adopted in multithresholding by shrinking the region to be thresholded gradually [153, 309]. It is used to deal

with the situations where the intensities of foreground and background in the image overlap.

3.1.2 Region Growing

Region growing is also a commonly used operation in image segmentation. It works based on the assumption that the voxels belonging to the same structure are connected and tend to be homogeneous with respect to some features (e.g. intensity). Hence the segmentation is achieved by identifying the connected regions with homogeneous features. Region growing algorithms are usually used as a preprocessing step for other more advanced segmentation methods. Their performance mostly rely on manual interaction for seeding. The results could be easily biased by image artefacts like noise, intensity inhomogeneity and partial volume effect [219]. At the beginning of the algorithms, the initial seeds and the homogeneity criterion have to be defined. Then the neighbouring voxels are gradually merged into the region until the homogeneity criterion is violated. According to the scheme of seed selection, the algorithms are divided into seeded and unseeded region growing.

Seeded region growing plants the seeds manually. In the simplest case, where the voxels of the foreground are connected and their intensities are homogeneous, the planted seed is initialised as the growing region. Then the neighbouring voxels are added in if their intensities lie within a range specified by a coefficient, which defines the allowed deviation from the mean of the region. In this process, the mean and standard deviation of the region are updated. Apart from the homogeneity criterion, the region size can also be used to constrain the region growing [198]. Once the region size exceeds a threshold, the growing is adapted to a more restricted criterion. For images with multiple structures to segment, Adams seeded region growing [3] can be applied, in which multiple seeds are planted and the new 'boundary voxel' is assigned to one of the regions according to the distance between the voxel and the region. For images with structure composed of unconnected regions, the grown region can be split by applying 'k-contraction', which removes the k voxels with the lowest intensities from the region [230].

Unseeded region growing chooses the seeds automatically. In an approach proposed in [169], one voxel is randomly selected as the seed. The 'boundary voxels' are merged into the existing region according to the distance measure. A new region is created if no more voxels can be found to be close enough to any existing region. The new region is seeded by an unlabelled voxel, which is closest to any existing region. The initial seeds can also be determined by applying thresholding techniques

as in [231]. The region growing is then controlled by the homogeneity parameter obtained from optimising an assessment function.

3.1.3 Deformable Models

Deformable models including deformable curves, surfaces and solid models have been widely applied in computer vision and computer graphics. Their potential has been proven in applications of medical image analysis such as segmentation, registration and motion tracking [191]. Deformable models incorporate both bottom-up image features and top-down prior knowledge (e.g. size, shape and location) of the structures [191]. During the process of segmentation, the initial deformable model evolves under the influence of driving forces. The topology of deforming model is constrained based on the prior knowledge of the structure geometry. The model evolves closely and eventually reaches the boundary of the structure to segment by interacting with the image. Deformable models are capable of taking structure variability into account and are flexible in manual intervention [191]. According to the mechanism of the models, they are classified into snakes, level sets and geodesic active contours.

Snakes are originated from the algorithm proposed in [145]. The model deformation is driven by the internal force derived from the geometry properties of the model and the external force derived from interacting with the image. The internal force maintains the regularity of the model while the external force attracts the model to the desired boundary of structure. The model evolution is completed when the minimum of the energy functional composed of the driving forces is achieved. In the original snake, the internal force is calculated from the curvature of the contour and the external force is defined by the gradient magnitude. The limitations of this method lie in several aspects. First, manual intervention is required for the initial state of the contour and the selection of deformation parameters. Second, the deformation might be trapped by local minima of the energy functional or stop when the gradient magnitude becomes too small. Third, the topology of the contour cannot adapt to the boundary in the image during the deformation process [176]. Numerous variants have been developed to overcome these shortcomings by improving or incorporating more constraints into the energy functional. Balloon snakes [67] use an inflation force, which allows the contour to pass over week edges and stop when the edge is strong. It makes snakes less sensitive to initialisation and achieve more stable performance. Topological constraints were proposed to enable the topology of contours to change conveniently during the deformation process [192, 85]. Several physical constraints regarding the metric, shape and resolution of the contours were also used in [85] to maintain the geometry

of the contours. Statistical techniques can be applied to extract the shape of models and define the range of parameters from training images. Such information can be incorporated into snakes as a priori knowledge to constrain the model deformation in testing images. The model evolves towards the objective boundary by interacting with the image features such as edges in active shape model [70, 71] or texture in active appearance model [72].

Snakes can generate closed curves and surfaces directly from the deformable models. Various constraints described above make the models robust to noise and spurious edges [219]. However, their applications are restricted by the computational complexity resulted from parametrisation of the contour, topology controlling and contour vertices redistribution [176].

Level sets are algorithms based on the theory firstly proposed in [208]. The deforming contour is implicitly embedded into a higher dimensional level set function with the contour as its zero level set. Therefore, instead of tracking the deforming contour composed of discrete points, segmentation progresses by tracking the zero level set. During this process, the topological changes are handled naturally. The model deformation of level sets is driven by the speed function, which is similar to the energy functional in snakes. The curvature component in the speed function maintains the smoothness of the contour. The other speed component calculated using image features attracts the contour to the desired boundary [127]. The commonly used image feature is the gradient as in [182]. However, the stopping criterion has to be carefully selected to avoid leakage and shrinkage of the deforming contour. The edge strength and weighted area flow have been incorporated into the speed function to improve the performance [148, 259].

Geodesic active contours combine level set formulation with contour energy functional minimisation. This implementation avoids the contour leakage, facilitates the algorithm computation and efficiently handle topological changes of the contour. To improve the accuracy and efficiency of segmentation, various a priori knowledge can be incorporated into the energy functional. For example, the a priori shape knowledge and statistical techniques were used for postprosessing in [165]; the gradient vector flow field, intensity distributions and distance information were incorporated to eliminate the dependence on initialisation and improve the segmentation accuracy in [212]. In addition, region based image features such as colour, texture, shape or area were also used to improve the model [312, 259]. Incorporation of a priori knowledge and statistical techniques also makes the deformable models more automatic and more robust to noise [176].

3.1.4 Machine Learning Techniques

Machine learning algorithms have been widely used in computer vision and image processing applications. In recent years, impressive results have been achieved by machine learning technologies. Based on features extracted from the image, the voxels are either grouped into several clusters directly or assigned predefined labels by classifiers trained using the same features. According to the necessity of a training process, we categorise the machine learning algorithms into unsupervised and supervised approaches.

Unsupervised clustering techniques group image voxels into a number of clusters automatically. The most commonly used clustering algorithms in medical image analysis include k-means (KM), fuzzy c-means (FCM) and Gaussian mixture models (GMM). The number of clusters has to be prespecified. In KM, the centres of clusters are randomly selected and the voxels are classified according to the distance between the voxels and the cluster centres [177]. In each iteration, the cluster centres are updated and the distance is recalculated. The segmentation is completed when the stopping criterion is satisfied (e.g. the cluster centres show little changes between iterations). Instead of performing hard segmentation, FCM calculates the probabilities of each voxel belonging to the clusters by introducing a membership function [37]. Several variants of FCM, such as bias field corrected FCM [5], enhanced FCM [276] and fast generalised FCM [54], have been proposed to improve the segmentation accuracy and efficiency. GMM estimates the probability density function of each cluster as a Gaussian distribution [228]. KM is usually applied to initialise the parameters of GMM. Then the segmentation and parameter updating are interleaved in an expectation maximisation (EM) approach until the stopping criterion is satisfied. These clustering algorithms are sensitive to initialisation. The segmentation could be easily biased if it only relies on image intensity. Spatial information can be incorporated into the algorithms as in [5]. A Markov random field (MRF) model [332] is usually combined with GMM to take the neighbouring information into account in an EM framework.

Supervised classification techniques achieve image segmentation by training classifiers using features extracted from labelled samples. The same feature space is constructed for the testing image and put into the classifier. Segmentation using classifiers requires a reasonable number of training samples and the training process usually takes a long time. However, once the classifier has been trained, the segmentation of testing images can be accurate and efficient. A variety of classifiers have been developed and applied in segmentation of medical images.

k-nearest neighbours (kNN) is the simplest classifier, which classifies each object (e.g., image, voxel) into the same class as the majority of its closest k training samples. The distance between each pair of testing and training samples is calculated based on their feature space. kNN can be either used alone or incorporated into other models in medical image classification [73, 293].

Support vector machine (SVM) was proposed in [41]. It was developed from computational learning theory. Kernel functions are used to map image features into a high dimensional space. Nonlinear functions (the hyperplanes) are learned to separate the objects into two classes by maximising the margin between their feature spaces. Finding the optimal hyperplane is a constrained optimisation problem, which can be solved by optimisation techniques like Lagrangian. The theory behind SVM guarantees its good classification performance. Because of this, the algorithm is not affected by local minima or suffers from high dimensional feature space. SVM has been successfully applied in applications of medical image analysis such as brain tumour segmentation [31] and tissue classification [199]. The disadvantages of SVM lie in the choice of kernel functions and the algorithm is not straightforward for multiclass segmentation. In practical uses, its application is also limited by the high computational complexity and extensive memory requirement in large-scale tasks [275].

Random forests (RF) algorithms are popular and accurate classifiers introduced in [47]. The algorithms can deal with both classification and regression problems. RF is an ensemble of decision trees. Each tree is trained independently based on a bootstrap sample chosen at random with replacement from the training set. Each bootstrap sample has the same size as the original training set. At each node of a tree, a random subset of features are selected to find the best split. The size of the feature subset is fixed for all the nodes and all the trees (typically $m = \sqrt{p}$, where p represents the total number of features and m is the size of subset). The trees keep growing until the maximum depth is reached. During this training process, the classification error of RF is internally estimated using out-of-bag data. In the testing phase, the feature space of a testing sample goes through every trained tree in RF and the new sample is assigned to the class which receives the majority of votes from all the decision trees. RF can handle large datasets with mixed features (both qualitative and quantitative) without making assumptions of data distribution. Redundant features can be easily ignored and missing data can be handled elegantly. It achieves competitive performance among machine learning methods in terms of accuracy and stability. For example, in the MICCAI grand challenge on MR brain image segmentation, the classifier using RF was ranked the third best among other techniques [307]. The main disadvantage of RF is that a large number of trees may slow down the algorithm in

real-time classification. And the algorithm might lead to overfitting for datasets with noisy classification tasks.

Bayesian network classifiers have shown remarkable classification performance among machine learning techniques. Given the class label, the classifiers learn the conditional probability distribution of each feature from the training set. Given a set of features, the probability of a class label assigned to a testing sample is computed by applying Bayes' rule. Then the class with the highest posterior probability is predicted for the testing sample [112]. Bayesian networks are directed acyclic graphs, where the vertices represent the random variables (i.e. the features and classes) and the edges represent direct correlations between them. Each variable is only dependent on its descendants given the state of its parents. Their correlations are represented by a local conditional distribution with probabilistic parameters. The joint probability distribution over all the variables is uniquely determined by combing these local conditional distributions [112]. The objective of Bayesian network classifiers is to find a network that describes the probability distribution over the training set effectively and efficiently. Usually this is achieved by using heuristic search over the space of all possible networks. A scoring function such as the Bayesian scoring function [69] or the function based on minimal description length [161] is adopted to evaluate the merit of each candidate network. Once the optimal network is determined, given the feature vector of a testing sample, the classifier outputs the class label that maximises the posterior probability. Naive Bayesian classifier is the simplified form of Bayesian networks. It makes strong independence assumption that all the features are conditionally independent given the class label. It performs well in classification tasks. However, it has been proven that the classifiers with less restrictive independence assumptions can perform even better [112]. Bayes' theorem and Bayesian networks have been successfully applied to applications of medical image analysis [303, 286]. Bayesian networks provide natural ways to structure data and yield intuitive predictions for new cases. They can readily handle incomplete data set and causal correlations. Prior knowledge is easily incorporated and much more difficult models can be estimated. The problem of overfitting is efficiently prevented. The disadvantages of Bayesian networks lie in the intensive computation, lack of defence of the priors and the assumption of the data distribution.

Artificial neural network (ANNs) are biologically inspired technologies, which simulate the way of processing information in the human brain [4]. The first conceptual model of an ANN was proposed in [190]. Following their work, numerous improvements and innovations have been developed. Classic ANNs consist of the input layer, the output layer and the hidden layers between them. Each layer contains

a number of neurons, which are fully connected with neurons at its adjacent layers. Each neuron has its input, transfer function and output. The input layer receives the data and the output layer provides the response of ANN to the input data. The hidden layers play a crucial part in processing the information. Once the structure of an ANN (i.e. the number of layers and the number of neurons at each layer) is determined, its performance depends on the transfer function of the neurons, the connection formula and the learning rule [4]. All the signals arriving at a neuron are firstly weighted, and then combined together to pass through a transfer function. The output for that neuron is then calculated. The most commonly used transfer function is the sigmoid function. The connections between neurons can affect the operation of an ANN by stimulation and inhibition mechanisms. Based on whether the feedback connection is available from the output of one layer to its previous layer or the same layer, feedforward and feedback architectures of ANNs are identified. The most commonly used learning rule of ANNs is the back-propagation rule. During the training phase, input data in the training set is fed forward through an ANN to compute the weights between neurons. The weights are optimised by backward propagation of the error iteratively until the specified accuracy is reached. Then all the weights are saved for the classifier. In the testing phase, the testing data is forward propagated through the network and its class label is predicted. ANNs are data driven, which makes them very flexible and powerful. They can create their own representations of the training data and the computation can be carried out in parallel. They can perform complex tasks without understanding the specifications of them. ANNs are well suited for dealing with large dataset, discovering regularities within a set of patterns and working out the relationships between variables that are difficult to describe with conventional algorithms [95]. On the other hand, ANNs are criticised as 'black boxes', computationally expensive and prone to overfitting [279].

Deep neural networks (DNNs) or deep learning (DL) algorithms have drawn great attention and achieved impressive classification results in computer vision in recent years. These techniques are inspired from the way the neocortex processes information received from the environment. Instead of preprocessing the received signals, the neocortex propagates the original data through a complex hierarchy of modules, which represents the data by finding exhibited regularities [14]. DL is defined as a subfield of machine learning which models the complex relationships among data by applying many layers of nonlinear information processing [86]. Each level extracts its distinct set of features, where high-level features are defined from lower-level ones and the same set of lower-level features can define different sets of higher-level ones [86]. This deep architecture represents the input data by learning

high-order features, which can be used for pattern analysis and classification by applying relatively simple classifiers described before. The popularity of DL arises from its ability to exploit complex relationships among data of large size and extract hierarchical features using both labelled and unlabelled data [86]. Depending on the objective of DL architectures, the algorithms are categorised into discriminative and generative deep networks. Discriminative deep networks aim to classify the data based on the high-order features extracted from the hierarchical data representation network. The labelled dataset is needed to train the deep network. Convolutional Neural Networks (CNNs) are typical discriminative deep networks consisting of several stages of convolution and subsampling processes [14]. As ANNs, CNNs are trained in a feedforward back propagation manner. At each stage, the input data (image for the first stage and feature map for the later stages) is convolved with a filter and then adds a bias to generate the convolution layer. The subsampling process is then applied, which consists of successive operations of summing, scalar weighting, adding bias and passing through a sigmoid function to the result from the convolution step [14]. This process results in a feature map, which is used as input data in the next stage. All the parameters in each stage including the filter coefficients, weights and biases are trainable. The convolution and subsampling steps in each stage of CNN produce features with invariance to shift, scale, rotation and translation. At the final stage, all the feature maps generated at the highest layer are rasterized and put into a conventional neural network to produce the final classification result. Owing to the ability to model the spatial relationships with hierarchical layers, CNNs are well suited for feature extraction and image processing applications. Generative deep networks extract high-order features from the unlabelled data for pattern analysis. A joint probability distribution over the observed data and its label is provided. The probability of observations given the class label can also be estimated. Deep Belief Networks (DBNs) are probabilistic generative deep networks composed of multiple layers of stochastic latent variables [86]. Compared with the traditional back propagation neural networks with deep layers, DBNs are computationally efficient. They select the parameters accurately and have no requirement of the labelled training dataset [14]. DBNs consist of several layers of Restricted Boltzmann Machine (RBM) stacked on top of each other. RBM restricts the neural network to only one visible layer and one hidden layer and only the connections between the layers are established. The layers of a DBN are connected by top-down generative weights, which are learned in a layer-by-layer manner by applying contrastive divergence algorithms proposed in [132]. For each RBM, the hidden layer abstracts the representation of the data at the visible layer and the new representation is served as input to the next higher level

RBM for more abstract representation. Each RBM stacked on the DBN strengthens the representation power of the network. The top two layers of a DBN have undirected, symmetric connections between them, which enables the top layer to associate with its memory contents from the lower layers [14]. If labelled data are available, after the pretraining described above, DBN can be fine tuned and turned into a discriminative network by learning bottom-up recognition weights through back propagation. This results in better discriminative performance and faster convergence compared with traditional feedforward back propagation neural networks [133]. DNNs have been successfully applied to a variety of applications such as face identification [274], speech detection [131], document analysis [68] and image classification [156]. In medical image analysis, they have also been used for high-level feature extraction and image segmentation [273, 65]. The main advantage of DNNs over shallow machine learning algorithms is learning hierarchical features inherently. The hidden layers can be repurposed from one application to another. However, it is difficult to optimise the parameters in all layers of networks. Moreover, DNNs perform as poorly as shallow machine learning techniques in some applications such as natural language processing and speaker identification. Domain knowledge like invariance of the features and regularisation of parameter constraint has to be determined for successful application of DNNs to a specific task. Considerable skill and experience is required for setting hyper-parameters such as the learning rate schedule, the strength of the regulariser, the number of layers and the number of neurons per layer [86]. A solid theoretical foundation in many aspects of DNNs still has to be established.

3.1.5 Graph Theory Based Methods

Graph theory based segmentation methods have shown great potential in computer vision applications. For these methods, the image elements are mapped onto a graph and the segmentation is achieved by partitioning the graph into several subgraphs such that each subgraph represents a segmented object. Let $G = (V, E)$ represents an undirected graph with a set of vertices V corresponding to the image voxels or regions. E stands for the edges connecting pairs of vertices in V. Each edge in E is given a weight w which defines the distance between two connected elements based on the local features (e.g. intensity, colour, location or texture). The image segmentation problem is formulated into partitioning the graph G into mutually exclusive components, such that each component is a connected subgraph $G^{'} = (V^{'}, E^{'})$, where $V^{'} \in V$, $E^{'} \in E$ and edges in $E^{'}$ only connect vertices in $V^{'}$ [215]. The segmentation is performed based on the criterion that the image elements represented by a subgraph $G^{'}$ should have

similar image features and the weights for edges in $G^{'}$ should be relatively low [107]. Formulating image elements into a graph enables the segmentation problem to be solved using graph theory tools. According to the segmentation principle, we divide the graph theory based algorithms into three categories including minimal spanning trees, graph cuts and shortest path based approach.

Minimal spanning tree based methods segment the image by finding the minimal spanning tree of the graph and breaking the edges with the highest weights [328]. The methods rely on the similarity and proximity of the graph vertices [215]. The minimal spanning tree guarantees the most similar image elements are connected in the graph. By cutting the edges with the highest weights, the graph is split into separate components, where the neighbouring subgraphs have the largest distance. Some improvements have been made to either speed up the implementation by adopting a recursive scheme [201, 159] or enhance the segmentation accuracy by taking the variation inside of the subgraph into account [107]. Minimal spanning tree based methods intuitively group image elements with similar features. They assume the elements in the same partition share the same labels. However, this assumption can be easily violated in dealing with real world images. Hence these methods are often used for preliminary segmentation of more advanced algorithms.

Graph cut theory was first used for image segmentation in [322]. A cost function is defined by the weights of graph edges and the segmentation is performed by globally minimising the cost function (minimal cut). The minimal cut method tends to segment small components in the images. To eliminate the bias, various normalised cut methods have been developed which incorporate information with regard to the size of segments [258], features of the interior region or the cut boundary [75, 310, 311]. Graph cut methods can also be combined with other segmentation techniques to further improve the performance. Typically, a MRF model is combined to incorporate the contextual constraints (e.g. the regularisation of image smoothness, shape of the segments, human interaction, etc.) into the segmentation approach. The segmentation problem is reformulated in a maximum a posterior (MAP)-MRF framework and solved by minimising an energy functional, which can be either globally optimised by applying combinatorial min-cut/max-flow algorithms [42] for binary segmentation or approximately optimised using α-expansion-move and $\alpha\beta$-swap-move algorithms [43] in multilabelling tasks. Graph cut methods establish well-defined mathematical relationships between the image elements and make the problem formulation more flexible. Efficient approximations have to be studied since finding a cut in an arbitrary graph or minimising an energy functional might be NP-hard [215]. The

solid theoretical foundation of graph cut makes it applicable to many applications in image segmentation.

Shortest path based methods convert the problem of finding the best boundary which separates the image elements to finding the shortest path between two vertices on the boundary. Human interaction is required to seed the initial points on the boundary. Then the shortest path is found between them to best fit the object of interest. The boundary is represented as a sequence of oriented edges with their weights in the graph. Searching for the shortest paths over the whole graph is time consuming. A variety of improvements have been made to alleviate this limitation such as confining the searching space [104], adopting heuristic search methods [186], introducing restrictions to the boundary [220] or incorporating prior knowledge of the graph [105]. Extension of shortest path based methods to 3D image segmentation is not straightforward and alternatively, searching minimal surface in 3D space can solve this problem fairly well [119]. Involving user interaction in shortest path based methods makes the segmentation effective and reliable. On the other hand, it might increase the human effort when the image is noisy, in which case a great amount of user guidance is required. This group of methods is more suitable for segmenting complex objects with explicit boundaries compared with other graph theory based techniques [215].

3.1.6 Atlas Based Methods

Atlas based methods are specific to medical image, in particular brain image segmentation. Due to the artefacts existing in brain images (as described in Section 2.3), the complexity of internal structures and the variability of brains among population, the anatomical a priori information is essential for guiding the segmentation approach. Prior information can be encoded as a set of predefined rules based on the property of brain structures or a template constructed from a set of manual annotations (i.e. atlas) [51]. An atlas is defined as two volumes: one intensity volume and one label volume. The label volume is the segmentation of its corresponding intensity volume. Generally, there are two types of atlases: topological and probabilistic atlases. A topological or deterministic atlas is a brain volume taken from a single subject, which is representative of a whole dataset in terms of image features like size, shape or intensity [51]. A group of manual annotations can be registered to a template or a common space (e.g. the Talairach space) and the deformed label volumes are averaged to construct a probabilistic atlas, in which the proportions of each voxel belonging to various structures are calculated. Compared with a topological atlas, a probabilistic

atlas is able to account for population variability. In the segmentation process, the registration between the atlas and the target image needs to be performed to propagate the atlas prior to the target image. Hence the segmentation problem is converted into a registration problem and the segmentation performance is dependent on the registration accuracy. In practice, rigid or affine registration is usually performed before applying nonrigid registration for further refinement. A multiresolution scheme is often adopted for algorithm efficiency and segmentation accuracy. Incorporation of atlases benefits the image segmentation especially in the cases where the shape information plays an important role. Based on the segmentation strategy used, atlas based methods can be categorised into label propagation, multi-atlas segmentation and probabilistic atlas based segmentation techniques.

Label propagation uses the topological atlas to incorporate the prior knowledge into the segmentation approach. The segmentation process is straightforward and relatively simple. By registering the atlas intensity volume to the target image, a transformation between them is obtained. Then the segmentation of the target image is achieved by applying the transformation to the atlas label volume. Interpolation methods such as nearest neighbour, linear, bilinear or cubic interpolation are used during the process. Label propagation is effective when the atlas image is similar enough to the target image. Therefore, it is more suitable for intra-subject applications such as longitudinal sequence analysis or multi-modality image fusion. The anatomical variability is not taken into account and the segmentation can be affected by the registration error.

Multi-atlas segmentation overcomes the shortcomings of label propagation by registering multiple atlases to the target image. This approach accounts better for anatomical variability and the effect of the registration error is considerably reduced. Two important factors are involved in the multi-atlas segmentation: atlas selection and label fusion. Atlas selection intends to pick out the most relevant atlases to avoid the bias introduced by the irrelevant atlases and reduce the amount of registration. The selection is performed based on meta-information (e.g. age) or image similarity (e.g. squared intensity difference or mutual information). It has been shown that using a subset of atlases with high image similarity results in improved segmentation accuracy compared with using the same number of random atlases, while the registration effort is reduced [6]. Once the atlases are selected, a label fusion strategy has to be applied to infer the final label map for the target image from all the candidate segmentations. The label fusion methods vary from the global to local strategies. The most commonly used global strategies include majority voting and global weighted voting based on normalised cross-correlation, mutual information or mean squared distance [16].

Local label fusion strategies include local weighted voting based on normalised cross-correlation or mutual information [16]. Recently, a probabilistic patch-based label fusion model was proposed, which models both the intensity likelihood and the label likelihood in a Bayesian framework [25]. The model maps each target voxel to a specific patch instead of a candidate voxel in each atlas, which leads to the improved segmentation performance. Generally, local label fusion strategies outperform global ones in segmenting high-contrast structures, while global strategies are less sensitive to noise in dealing with low-contrast neighbouring structures [16]. Besides, to alleviate the burden of registration, a compromised approach can be applied, in which all the atlases are first registered to the target image by affine registration, then after atlas selection only the selected atlases are nonrigidly registered to the target image to perform the label fusion [25].

Probabilistic atlas based techniques make use of the probability values by registering the atlas to the target image. A single registration is performed and the probabilistic atlas accounts for the anatomical variability. A probabilistic atlas is often used to encode the spatial information modelled in a Bayesian framework [293]. The probabilities can also be used to initialise the parameters of the probability density function of each class, which is usually modelled by a Gaussian mixture model [80]. Some approaches also extract context features from the probability maps and use them together with image appearance features like intensity or Haar-like features to train a classifier [307]. In addition, a subset of samples with high probabilities are also extracted from the atlas to train a classifier, estimate class distributions or initialise an active contour [51].

3.2 Evaluation of Image Segmentation

Performance evaluation is a non-trivial part of image segmentation algorithms. A few papers reviewed the evaluation methods for generic image segmentation in the literature [330, 58, 333]. The most commonly used approaches are subjective and objective comparisons. In subjective evaluation, the segmentation results are visually examined by observers. Due to the inherent subjectivity, this evaluation approach suffers from inter- and intra-variability. It is difficult to obtain an unbiased evaluation to assess the effectiveness of algorithms. In order to minimise the bias, a sufficiently large group of evaluators have to conduct intensive visual evaluation on the segmentation results of a sufficiently large set of test images [330]. The testing has to be conducted under well-defined guidelines to reduce the favouritism to specific algorithms or parameter

settings. Thus subjective evaluation is tedious and time-consuming and unsuitable in real-time tasks. In the experiments conducted in this thesis, we usually use it as a complement to objective evaluation to provide intuitive results and comparisons. In objective evaluation, the segmentation results are compared with a set of reference images, which are labelled manually and considered as ground truth. Manual delineation takes a great amount of time and effort, which could be improved by means of segmentation software. It also suffers from the intra- and inter-rater variablity, which could be mitigated by applying label fusion algorithms like STAPLE [313]. The segmentation is evaluated by the agreement between the segmentation results and the ground truth calculated using defined metrics. In the following, we will introduce the commonly used metrics in objective evaluation of segmentation methods and divide them into overlap measures, distance measures and probabilistic metrics.

3.2.1 Overlap Measures

Literally, overlap measures calculate the overlap between the segmentation result A and the ground truth B. Most of these metrics are defined based on true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), where TP and TN represent the numbers of voxels correctly classified as foreground and background, respectively; FP and FN stand for the numbers of voxels misclassified as foreground and background, respectively. A set of metrics can be defined as listed in Table 3.1.

The Receiver Operating Characteristic (ROC) curve, which represents the Sensitivity against $1-$Specificity with various decision thresholds, is also used to evaluate the segmentation performance. The accuracy is often measured by the area under the ROC curve (AUC), which can be calculated using trapezoidal integration [45]. Besides, instead of evaluating the voxel-wise agreement between the segmentation results and the ground truth, the object-wise agreement is also used to calculate some metrics like Dice similarity coefficient, Sensitivity, Specificity and Accuracy. These object-wise metrics can be used for evaluating specific segmentation approaches such as lesions segmentation, where the number of lesions rather than the voxels are counted.

3.2.2 Distance Measures

Distance measures calculate the distance between the surfaces of the segmentation result S_A and the ground truth S_B. S_A and S_B are two finite sets of points.

Table 3.1 Overlap Measures

Name	Equation
Dice similarity coefficient	$\frac{2TP}{FP+FN+2TP}$
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{FP+TN}$
Accuracy	$\frac{TP+TN}{TP+FP+FN+TN}$
Error rate	$\frac{FP+FN}{TP+FP+FN+TN}$
Over estimation fraction	$\frac{FP}{TN+FN}$
Under estimation fraction	$\frac{FN}{TN+FN}$
Similarity index	$\frac{2\lvert A\cap B\rvert}{\lvert A\rvert+\lvert B\rvert}$
Jaccard index	$\frac{\lvert A\cap B\rvert}{\lvert A\cup B\rvert}$
Precision	$\frac{\lvert A\cap B\rvert}{\lvert B\rvert}$
Recall	$\frac{\lvert A\cap B\rvert}{\lvert A\rvert}$
Relative absolute volume difference	$\frac{\lvert A\setminus B\rvert}{\lvert B\rvert}$

Note: $|A|$ represents the number of voxels in A; $\cap$, $\cup$ and $\setminus$ are set interaction, union and difference operators, respectively.

Hausdorff distance (HD) is the most commonly used distance measure, which is defined as

$$HD = max(h(S_A, S_B), h(S_B, S_A)), \tag{3.1}$$

where

$$h(S_A, S_B) = \max_{a \in S_A} \min_{b \in S_B} d(a,b), \tag{3.2}$$

and $d(a,b)$ calculates the distance, which can be Euclidean distance or squared Euclidean distance, between the points a and b. $h(S_A, S_B)$ is the directed Hausdorff distance from S_A to S_B. For any point $a \in S_A$, the minimum distance is found to its closest point $b \in S_B$. Then all the minimum distances for all the points in S_A to their closest points in S_B are ranked and the maximum is found which defines $h(S_A, S_B)$. The Hausdorff distance is the maximum of $h(S_A, S_B)$ and $h(S_B, S_A)$ which measures the degree of mismatch between two point sets [134].

Mean distance (MD) measures the mean distance of all points in S_A to their closest points in S_B, which can be formulated as

$$MD = \frac{\sum_{a \in S_A} \min_{b \in S_B} d(a,b)}{|S_A|}, \tag{3.3}$$

where $|S_A|$ stands for the number of points in S_A.

Mean average surface distance (MASD) averages the mean distance from both S_A to S_B and S_B to S_A. It is calculated by

$$MASD = \frac{1}{2}\left(\frac{\sum_{a\in S_A}\min_{b\in S_B} d(a,b)}{|S_A|} + \frac{\sum_{b\in S_B}\min_{a\in S_A} d(b,a)}{|S_B|}\right). \tag{3.4}$$

Average symmetric surface distance (ASSD) measures the mean distance for all points in both S_A and S_B to their corresponding closest points in S_B and S_A. It is defined as

$$ASSD = \frac{\sum_{a\in S_A}\sum_{b\in S_B}\min\, d(a,b) + \sum_{b\in S_B}\sum_{a\in S_A}\min\, d(b,a)}{|S_A|+|S_B|}. \tag{3.5}$$

3.2.3 Probabilistic Measures

The overlap and distance measures described above are all used for evaluation of hard segmentation. In fuzzy segmentation, where the probabilities of each voxel belonging to multiple classes are calculated, the probabilistic measures can be defined to evaluate the segmentation accuracy. For example, the probabilistic similarity index (PSI) is defined as

$$PSI = \frac{2\sum_{x\in A} P(A(x)=1|B(x)=1)}{\sum_{y\in B,B(y)=1} 1 + \sum_{x\in A} P(A(x)=1)}, \tag{3.6}$$

where $P(A(x)=1|B(x)=1)$ is the probability of voxel x in A correctly classified as the foreground. The first component in the denominator counts the number of foreground voxels in B and the second component sums up the probabilities of all the voxels classified as the foreground in A. The PSI is equivalent to the Dice similarity coefficient if the final segmentation is binary [172].

Similarly, the probabilistic overlap fraction (POF) and the probabilistic extra fraction (PEF) are defined as

$$POF = \frac{\sum_{x\in A} P(A(x)=1|B(x)=1)}{\sum_{y\in B,B(y)=1} 1}, \tag{3.7}$$

and

$$PEF = \frac{\sum_{x\in A} P(A(x)=1|B(x)=0)}{\sum_{y\in B,B(y)=1} 1}, \tag{3.8}$$

where $P(A(x) = 1|B(x) = 0)$ is the probability of voxel x in A misclassified as the foreground. The POF is equivalent to the Sensitivity in hard segmentation.

3.3 Registration Algorithms

Image registration is widely used in remote sensing, medical image analysis, cartography and computer vision [337]. Registration algorithms vary in different ways such as image dimensionality, registration basis, nature of transformation, optimisation strategy, image modalities, etc. [180]. In this section, the main procedures of image registration and the most popular methods used at each stage are reviewed.

3.3.1 Feature Matching

In order to find the transformation between the source and target images, the correspondences between their features have to be established first. Either the raw image voxels or the features extracted from the images can be used for this purpose.

Voxel property based methods operate directly on the image intensities. Reductive approaches (e.g. principal-axes and moment-based methods) are used to reduce the image grey values to a representative set of scalars and orientations, based on which the registration is performed. Alternatively, the full image content can be used throughout the registration process by applying correlation-like methods, Fourier methods or mutual information methods. The normalised cross-correlation (NCC) and its variants search for the corresponding window pairs by measuring the intensity similarities. NCC methods can account for image translation, slight rotation and scaling and even affine transformation. The sequential similarity detection algorithm (SSDA) is more efficient than NCC methods by exploiting a simpler distance measure and a sequential search strategy [29]. Fourier methods outperform correlation-like methods in dealing with images taken under varying conditions or corrupted by correlated and frequency-dependent noise, in which case the images are represented in the frequency domain [337]. The phase correlation method [158] aligns two images shifted relative to one another based on the Fourier Shift Theorem [44]. It calculates the cross-power spectrum phase of two images and then determines the peak location in its inverse transform. This method was combined with polar-log mapping of the spectral magnitude to account for additional rotation transform [84]. Mutual information methods are originated from information theory [178]. They perform registration between two images by maximizing their statistical dependency, which is measured by the joint

entropy, the mutual information or the normalised mutual information. These methods are well suited for multimodal registration.

Voxel property based methods are usually computationally efficient, easy for hardware implementation and can be used in real-time tasks. However, their applications are limited by the simple-shaped windows in matching two images with complex deformation. The 'remarkableness' of the window content might lead to mismatching due to its non-saliency [337].

Feature based methods use salient features detected from the images to establish the correspondences between two images. The features can be landmarks and segmentation based structures. Anatomical landmarks (i.e. salient points identified by the users) and geometrical landmarks (e.g. corners, line intersections, local curvature extrema, etc.) manually or automatically extracted from the images are mostly used to estimate rigid or affine transformations. They can also be used for more complex transformations if the landmarks are sufficient. Compared with the entire image content used in the voxel property based methods, the landmark set is sparse, which relatively speeds up the optimisation procedure. Segmentation methods can be applied to extract features like surfaces, curves and volumes, which are used in either rigid-model or deformable model based registration approaches. Segmentation based features can be easily and efficiently obtained. However, the registration accuracy can be affected by the segmentation performance [180].

Once the feature space is determined, the corresponding features in the source and target images have to be matched. Feature matching can be achieved by using spatial relations or invariant descriptors. Methods based on spatial relations exploit the distance between features and their spatial distributions, such as the graph matching algorithm [117], clustering technique [271], chamfer matching [30], relaxation technique [227] and the iterative closest point algorithm [36]. Feature descriptors invariant to image deformations can also be used to estimate the feature correspondence, such as the intensity similarity in a local neighbourhood represented by cross correlation [1] or correlation coefficients [335], shape descriptor [214].

3.3.2 Transformation Estimation

Based on the established feature correspondences, a transform model is estimated to overlap the source image with the target image. The model could be global or local according to the image domain where the transformation applies. The global model maps the entire images using the transform function defined by a set of parameters, while the local model performs piecewise mapping by the transform function with

varying parameters derived from the local image content. The commonly used models for the simplest to the most complex transformations are described in the following.

Bivariate polynomial transformations of low degrees can be used to represent rigid, affine and projective transformations. Rigid transformation is applied between images with deformation of translations and rotations. It can be determined by 6 parameters in 3D space. Affine transformation maps parallel lines to parallel lines, which also allows rescaling and shearing and can be defined by 12 parameters. Projective transformation maps lines to lines without preserving the parallelism. It is often used in 2D/3D registration. These three transformations can also be represented by a constant matrix. The second or the third order polynomial models are used when more complex deformation is required. The use of higher order polynomials is restricted to avoid the unnecessary warping in image regions away from the control points [337].

Radial basis functions are global mapping methods composed of a linear combination of basis functions. These models are motivated by the interpolation theory, where the displacements calculated at the control points are propagated to the rest of the image domain by applying interpolation methods. The displacement at each point only depends on its distance to the control points, not on its position or grey level value. These functions are able to deal with the locally varying geometric distortions. The basis functions that can be used include multiquadrics [242], Gaussian functions [12], clamped-plate splines [185], thin-plate splines [40], cubic B-splines [160], Wendland functions [315], cosine functions [260], etc. Thin-plate splines are the most often used basis functions, which minimise the quadratic variation functional of the bending energy [337]. This method achieves good registration. However, it exhibits a few shortcomings such as the inverse inconsistency problem, suppression of the local image warping, incapability to deal with landmark localisation error and demanding computation [268]. A number of approaches have been proposed to tackle these shortcomings [141, 167, 236, 92].

Elastic models approximate transformation between images. They are capable of mapping images with complex deformation. The images are modelled as elastic bodies, which are stretched into alignment under the external force defined by the image similarity and the internal force defined by stiffness or smoothness constraints [49]. The registration is achieved by finding the minimum energy state or equilibrium in an iterative and pyramid approach [337]. Fluid registration methods [48] using the viscous fluid model are capable to handle very localised deformation, where the target image is modelled as a thick fluid that flows out to match the source image. Other nonrigid registration methods include diffusion models [282], level sets [297], flows of diffeomorphisms [287].

3.3.3 Optimisation Methods

Optimisation methods aim to find the optimal transformation between two images. The mapping function usually composed of a similarity term and a regularisation term is defined by a set of parameters. For bivariate polynomial transformations, the parameters can be easily computed by minimising the sum of squared errors at the control points. For local transformations, it is also possible to compute the local displacements directly from the local image content such as in optical flow based approaches. Most often, the optimal parameters are searched by applying optimisation methods. According to the nature of the parameters and the objective function, the optimisation methods can be divided into continuous, discrete and heuristic and metaheuristic approaches [268]. Continuous optimisation methods solve problems where the parameters assume real values and the objective function is differentiable [268]. Various methods belonging to this category but differing in factors like the step size or the search direction include gradient descent [338], conjugate gradient [124], Powell's method [166], Quasi-Newton [223], Levenberg-Marquardt [17] and stochastic search methods [28]. On the contrary, discrete optimisation methods solve problems where the parameters take discrete values. This class of methods include graph based methods [277], belief propagation methods [129] and linear programming approaches [267]. The continuous and discrete optimisation methods are constrained by the objective functions they can optimise. Conversely, the heuristic and metaheuristic methods are able to deal with a wide range of problems in a large search space without guarantee regarding to the optimality of the solution [268]. For example, greedy approaches [254] and evolution algorithms all belong to this group [247].

3.3.4 Image Resampling

The last step of a registration approach is image resampling, which applies the transformation derived from the previous steps to the source image. This process can be implemented in a forward or backward manner. The forward method directly transforms the source image using the estimated mapping function to find the corresponding coordinates in the registered image, while the backward method applies an inverse transformation to the registered image coordinates to find the corresponding positions in the source image. In practice, the backward approach is often chosen in order to avoid holes or overlaps in the registered image resulted from discretisation and rounding [337]. The interpolation is usually achieved by convolving the image with an interpolation kernel. To reduce the computation cost, multidimensional convolution

is usually decomposed to separate one dimensional convolutions. The optimal interpolation kernel, the sinc function, is impractical to calculate, thus various interpolation methods have been proposed to approximate the sinc kernel. The commonly used interpolation methods include the nearest neighbour, linear and cubic functions, quadratic splines [90], cubic B-splines [163], Gaussians [122], truncated sinc functions [280], etc. Nearest neighbour interpolation is suitable for images with a low number of grey levels, in which case the introduction of synthetic intensities is undesired. Linear interpolation gains the most popularity because it offers the best tradeoff between the accuracy and computation complexity. For transformed images with significant geometric enlargement, the cubic interpolation is preferred. Higher order interpolation methods produce more accurate results at a cost of high computation effort.

3.4 Evaluation of Image Registration

Evaluation of image registration algorithms is challenging in several aspects: the registration performance could be affected by the errors introduced at every stage; it is hard to distinguish the registration errors from the actual physical differences of the images; the ground truth of the transformation is usually unavailable. Generally, the registration inaccuracies consist of three parts: landmark localisation error, feature matching error and image alignment error [337]. The landmark localisation error results from the intrinsic error of feature detection methods. It cannot be measured directly from the resulting images but from the mean precision of the feature detection algorithms applying on various image types [337]. An optimal feature detection algorithm is often chosen to mitigate the landmark localisation error. The feature matching error measures the mismatching of features when establishing the correspondences between them. In most cases it can be alleviated by robust matching algorithms. The false matches can be identified by a consistency check or cross validation [337]. The image alignment error measures the difference between the estimated transformation derived from the registration process and the actual between-image distortion. This error is always present in practice due to the lack of a priori knowledge about the image distortion and/or the imprecise estimation of the parameters representing the transformation [337]. The image alignment error can be measured qualitatively or quantitatively. In qualitative evaluation, the images are visually inspected by the observers using visualisation tools such as colour overlays or difference images [130]. The registration is judged as ‘success’ or ‘failure’ depending on whether it reaches the required application accuracy or not. However, this approach is not very effective

particularly for algorithms that generate a fairly uniform range of errors either below or above the expected accuracy [130]. A wide range of quantitative metrics have been proposed in the literature to measure the image alignment error. They can be divided into volume overlap measures, surface overlap measures, boundary distance error, volume similarity measures and transformation consistency measures, which are elaborated in the following.

3.4.1 Volume Overlap Measures

Volume overlap measures are specifically designed for evaluating brain image registration algorithms in [151]. The original brain images together with their structure labels form the dataset. The registration is performed between the source and target images. The obtained transformation is then applied to the source label image. The registration performance is evaluated by measuring the agreement between the registered label image S and the target label image T for each structure r. A set of region based metrics are defined as below.

Target overlap (TO) is a measure of sensitivity, which is calculated by

$$TO = \frac{\sum_r |S_r \cap T_r|}{\sum_r |T_r|}, \tag{3.9}$$

where $|S_r \cap T_r|$ represents the number of voxels in the intersection between the transformed source region S_r and the target region T_r; $|T_r|$ represents the number of voxels in the region r of the target image.

Mean overlap (MO) is a special case of the Dice similarity coefficient and defined as

$$MO = \frac{2\sum_r |S_r \cap T_r|}{\sum_r (|S_r| + |T_r|)}. \tag{3.10}$$

Union overlap (UO) is another form of the Jaccard coefficient

$$UO = \frac{\sum_r |S_r \cap T_r|}{\sum_r |S_r \cup T_r|}. \tag{3.11}$$

The false negative rate (FNR) and the are defined to measure the disagreement between the registered source label image and the target label image, which are computed by

$$FNR = \frac{\sum_r |T_r / S_r|}{\sum_r |T_r|} \tag{3.12}$$

and

$$FPR = \frac{\sum_r |S_r/T_r|}{\sum_r |S_r|}, \tag{3.13}$$

where $|T_r/S_r|$ indicates the number of voxels in T_r but not in S_r.

3.4.2 Surface Overlap Measures

Similar to the volume overlap measures, the surface overlap measures are also specifically designed for evaluating brain image registration algorithms [151]. They are stated as more robust to structure segmentation bias and imaging artefacts that affect the cortical thickness [151]. Apart from the processings described in Section 3.4.1, the surface of each target image is generated and then intersected with both the target and the transformed source label images. Then the metrics defined for the volumes can be used to measure the overlap agreement and error between the target label surface and the transformed source label surface.

3.4.3 Boundary Distance Error

Following the same registration procedure described in Section 3.4.1, the boundaries for each structure are extracted from both the target label image and the transformed source label image. Then the registration performance is evaluated by computing the distance between each pair of region boundaries. For each region r, the distance error (DE_r) between the source region boundary S_rB and the target region boundary T_rB is calculated by

$$DE_r = \frac{\sum\limits_{s \in S_rB} \min\limits_{t \in T_rB} d(s,t)}{|S_rB|}, \tag{3.14}$$

where $d(s,t)$ measures the distance between the point s on the transformed source region boundary S_rB and the point t on the target region boundary T_rB; $|S_rB|$ counts the total number of points that compose S_rB. DE_r measures the mean distance of all points on S_rB to their closest points on T_rB.

A modified Hausdorff distance is also defined to measure the distance between the transformed image boundary and the target image boundary:

$$DE = \frac{1}{2}\sum_r \left(\frac{\sum\limits_{s \in S_rB} \min\limits_{t \in T_rB} d(s,t)}{|S_rB|} + \frac{\sum\limits_{t \in T_rB} \min\limits_{s \in S_rB} d(t,s)}{|T_rB|} \right). \tag{3.15}$$

3.4.4 Volume Similarity Measures

Volume similarity measures evaluate the registration accuracy by either calculating the differences between the transformed source label image S and the target label image T or directly measuring the similarity between the transformed source image and the target image. In the former case, the volume similarity (VS) can be defined as

$$VS = 2\frac{\sum_r(|S_r| - |T_r|)}{\sum_r(|S_r| + |T_r|)}. \tag{3.16}$$

The image based similarity measures such as root of mean squares (RMS), NCC and normalised mutual information (NMI) and are widely used to evaluate image registration algorithms. For brain image registration, the tissue overlap measures such as the Jaccard index and Dice similarity coefficient defined based on the tissues are also frequently used. The performance of registration algorithms can also be evaluated by registering a group of source images to a target image and then measuring the sharpness of the average image or the magnitude of the standard deviation image [324, 64]. However, all these metrics are unreliable in distinguishing reasonable from inaccurate registrations as stated in [234].

3.4.5 Transformation Consistency Measures

Besides comparing the transformed source image with the target image, the registration performance can also be evaluated by measuring the transformation consistency such as the inverse consistency metric and the transitivity metric.

For registration between the source image A and the target image B, the inverse consistency metric measures the inverse consistency error (ICE) between the forward transformation $\mathbf{T}_{A\to B}$ and the backward transformation $\mathbf{T}_{B\to A}$, which is defined as

$$ICE = \sum_{\mathbf{x}\in A} \|\mathbf{T}_{B\to A}(\mathbf{T}_{A\to B}(\mathbf{x})) - \mathbf{x}\|^2, \tag{3.17}$$

where $\mathbf{x}$ denotes the coordinates of each voxel in A and $\|\cdot\|$ is the standard Euclidean norm.

The transitivity metric measures how well the pairwise registrations performed on a group of images satisfy the transitivity property [64]. Given three images A, B and C, three transformations $\mathbf{T}_{A\to B}$, $\mathbf{T}_{B\to C}$ and $\mathbf{T}_{C\to A}$ are obtained by performing pairwise registrations. Ideally, the three transformations should project one point from image A to B to C and then back to the original position in A. The transitivity error (TE) with respect to image A is defined as

$$TE = \sum_{\mathbf{x} \in A} \|\mathbf{T}_{C \to A}(\mathbf{T}_{B \to C}(\mathbf{T}_{A \to B}(\mathbf{x}))) - \mathbf{x}\|^2. \tag{3.18}$$

As stated in [64], both ICE and TE are necessary but not sufficient evaluation metrics for evaluating image registration performance.

3.5 Statistical Analysis Methods

For the experiments in the thesis, we use a paired-sample t-test [15] to analyse the impacts of the parameter settings on the performance of specific methods. Repeated-measures Multivariate Analysis of Variance (MANOVA) [189] is applied to compare the performance of different methods with regard to several metrics on the same database.

Paired-sample t-test. A paired-sample t-test measures the difference between two sets of results with the significance level usually set at 0.05. The two sets of results are produced by the same method with different parameters and measured by the same metric on the same database. Given two sets of results for n subjects, the t-statistic is calculated by

$$t = \frac{\bar{d}}{\sqrt{s_d^2/n}}, \tag{3.19}$$

where $\bar{d}$ and s_d are the mean and standard deviation of the differences between two sets of results. Under the null hypothesis, t follows a t-distribution with $n-1$ degrees of freedom. By looking up the value of t in the t_{n-1} distribution table, the p-value for the paired sample t-test is obtained, which tells if a significant difference exists between the two sets of results.

Repeated-measures MANOVA. Repeated-measures MANOVA works in several steps. At first, the multivariate analysis tells whether significant differences exist among the performance of different methods with respect to the combination of all metrics. If significant differences exist, the univariate analyses are undertaken to explore whether the significant differences exist among the methods in terms of each metric. Then for each metric where significant differences exist, the Bonferroni *post hoc* test is utilised to conduct all the pairwise comparisons to determine which pairs of the results are significantly different with the significance level usually set at 0.05. Finally, based on all the pairwise comparison results with respect to each metric, the performance of different methods is compared taking all metrics into account.

3.6 Summary

This chapter presented a technical literature review on generic image segmentation and registration approaches, the measures commonly used to evaluate their performance and the statistical analysis methods for comparing the performance of various approaches. For image segmentation, methods have been reviewed from basic techniques such as thresholding and region growing to more advanced approaches like deformable models and machine learning techniques. Apart from the segmentation methods for general images, atlas based methods, which are specific for medical image segmentation, have also been described. For each category of methods, their principles, advantages, disadvantages and some applications in medical image analysis have been provided. Image segmentation performance can be evaluated using overlap measures, distance measures and probabilistic measures. For image registration, the main procedures including feature matching, transformation estimation, parameter optimisation and image resampling have been elaborated. The general methods used in each stage have also been reviewed. Image registration accuracy can be evaluated by volume overlap measures, surface overlap measures, boundary distance error, volume similarity measures and transformation consistency measures. The metrics based on local regions are more reliable than those defined on the whole volume or the transformations. The statistical analysis methods including the paired-sample t-test and MANOVA, which are used for comparing the performance of different approaches, have also been described.

Chapter 4

Brain Segmentation

4.1 Introduction

Brain segmentation, also known as skull stripping, is an important preprocessing step for brain image analysis. Manual delineation of the brain surface is likely to achieve more accurate results than automatic methods for brain segmentation, however, it is labour intensive and time-consuming, usually taking between 15 minutes and 2 hours per volume [263]. Also, experts need sufficient training to obtain less subjective segmentation results. Furthermore, intra- and inter-rater variability may occur during segmentation [313]. Therefore, fast and robust automatic brain segmentation methods are desired.

4.1.1 Previous Work

A range of state-of-the-art brain segmentation methods have been proposed [263, 252, 249, 336, 101, 213, 57]. However, the problem is still not considered to be solved [137].

Thresholding, region growing and morphology based methods are commonly used for skull stripping. An algorithm using automated thresholding and morphological operations was proposed in [164] for brain segmentation. A region growing method was applied in [213] for skull stripping, where histogram analysis and morphological operations were used as preprocessing approaches. The most representative morphology based method is the Brain Surface Extractor (BSE) [252], which applies a series of thresholding, morphological operations and edge detection to identify brain surfaces. First, an anisotropic diffusion filter is applied to smooth noisy regions in the image. Second, a Marr-Hildreth edge detector is used to identify important anatomical bound-

aries. Next, regions are eroded into components and the largest connected component is selected as the brain region. Then dilation is used to restore the brain. Finally, small pits in the surface and some holes within the volume are filled by a closing routine. The weakness of morphology based methods is that often the parameters need to be fine-tuned to overcome specific problems.

Deformable model based methods usually define a surface model and then fit it to the brain surface. Zhuang *et al.* [336] proposed a model-based level set to extract the brain surface. They used two terms separately derived from the curvature and the intensity information of the brain image to enforce the evolution of the zero level curve. Two most popular deformable model based brain segmentation methods are the Brain Extraction Tool (BET) [263] and the Hybrid Watershed Algorithm (HWA) [249]. The BET evolves a tessellated mesh of triangles to fit the brain surface. Three constraints are used to drive the fitting: the first one controls the smoothness of the surface; the second one maintains the equal distribution of the vertices over the surface; the third one deforms the model to the correct brain surface. The HWA combines the watershed algorithm and deformable surface models. The watershed algorithm is performed to extract an initial brain with the assumption of connectivity of the white matter. Then, a deformable model is applied to refine the output, recovering falsely removed tissue using statistical atlas information. A new implicit deformable model, which is represented by a set of radial basis functions (RBFs), was proposed in [171]. Driven by the internal and external forces, the deformable model evolves from its initial state to the desired brain boundary by iteratively updating the RBFs. A registration step can also be incorporated into the deformable model for brain segmentation. A strategy proposed in [26] uses the level set method for brain segmentation and the initialisation of the brain surface is achieved through dense registration. Wang *et al.* [336] presented a skull stripping method for large-scale studies. Their approach includes the co-registration with an atlas and the refinement of the deformable surface based on the prior information obtained from a set of brain images. A method combining deformable surface models and a fuzzy tissue classification was proposed in [278]. In contrast to morphology based methods, the deformable model based approaches seem more robust and less sensitive to parameter settings. But the results often depend on the initialisation and it may lead to self-intersection in the process of surface evolution.

The robust, learning-based brain extraction system named ROBEX [137] combines a generative model with a discriminative model. The discriminative model is a voxel-based Random Forest trained to detect the brain boundary; the generative model ensures the brain surface is plausible using a point distribution model. The final segmentation is obtained by refining the contour using graph cuts. Graph cuts was

also used in [244] to remove dura in brain segmentation. In addition, a skull stripping technique for neonatal brain MR images was presented in [179], which incorporates prior shape information within a graph cut framework.

The segmentations of other brain images can also be used for brain segmentation through atlas based or patch based approaches. A consensus masking framework named McStrip was proposed in [229], which incorporates atlas based segmentation, intensity thresholding and edge based masking. A multi-atlas based skull stripping method was presented in [93]. A patch based approach BEaST was used in [102] for brain segmentation in a multi-resolution framework.

Among all the brain segmentation methods, BET is widely used due to its simplicity, accuracy and insensitivity to parameter settings. However, some inherent weak points exist. Firstly, typically a large number of iterations (typically $1,000$) are needed to find an acceptable surface, which directly results in a long processing time needed to segment a brain volume. Secondly, it lacks an efficient strategy to deal with the surface self-intersections, which might lead to inaccurate segmentation in some regions of the brain.

4.1.2 Our Work

We present an improved BET method in this chapter. It substantially decreases the number of iterations by enhancing vertex displacement combined with an independent surface reconstruction step at each iteration. Apart from the search path used in the original BET to calculate the local intensity threshold, we add one more search path at the bottom of the brain to better reflect the neighbouring intensity information. Thus the surface is more likely to evolve more correctly. When the optimal brain surface is found, a false positive reduction process based on FCM [55] is performed to further refine the segmentation results. The performance of the improved BET method is compared with that of three other methods mentioned in the previous section (the original BET, BSE and HWA) using several metrics and datasets. BSE and HWA are commonly used brain segmentation techniques. They are embedded in publicly available softwares (BrainSuite and FreeSurfer), which are also widely used in brain image analysis. It is straightforward and convenient to produce the brain segmentation results using the existing software and compare them with our results. Each of the three comparison methods represents a different brain segmentation approach: deformable model based, morphology based and hybrid approach, respectively.

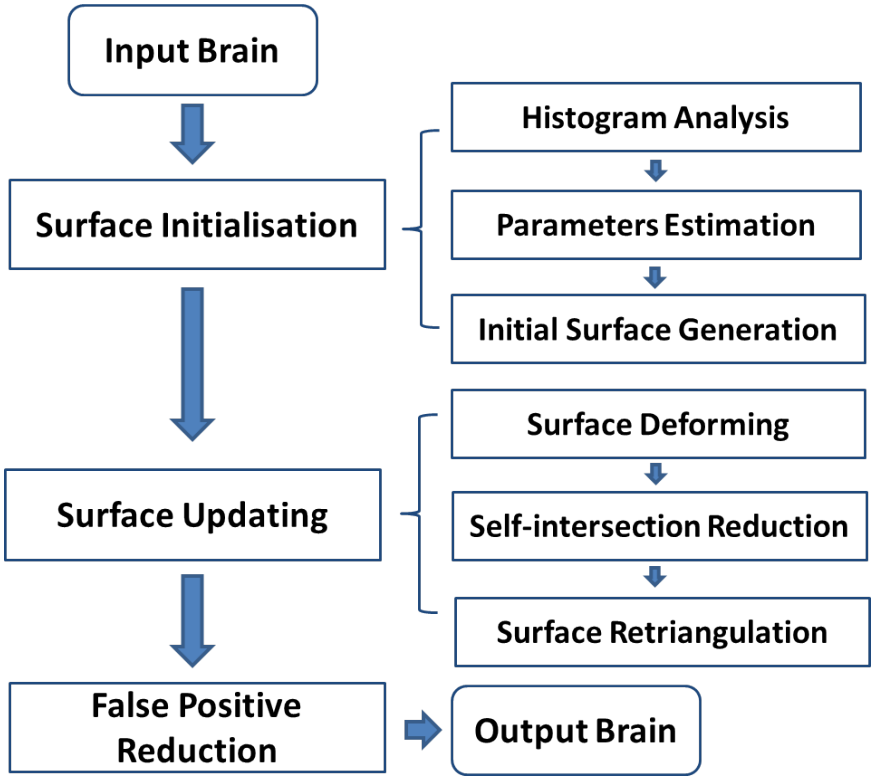


Fig. 4.1 Workflow of the method.

4.2 Method

A triangular tessellation of an ellipsoidal surface is initialised inside the brain as the deformable model. This model evolves following the force derived from the local intensity information. An independent surface reconstruction process is applied at the end of each surface update. Empirically, the surface converges within about 50 iterations and the optimal brain surface is obtained. Finally, we use FCM and some morphological operations to remove the non-tissue regions from the previously extracted brain. The entire workflow is shown in Fig. 4.1.

4.2.1 Surface Model Initialisation

An intensity histogram is used to find the robust global minimum and maximum intensities of the brain image. The minimum intensity, referred to as t_2, is the intensity below which lies 2% of the cumulative histogram. The maximum intensity t_{98} is found in a similar way. The global threshold t, which distinguishes the brain from the

background, is estimated as a 90%/10% weighted average of t_2 and t_{98}

$$t = 0.9t_2 + 0.1t_{98}. \tag{4.1}$$

The threshold t and maximum intensity t_{98} are used to estimate the centre of gravity (COG) of the brain. The bounding box for the voxels, whose intensities lie between t and t_{98}, can be generated. The centre of the bounding box is considered as the COG. The median intensity of all the voxels within the bounding box is found, which is referred to as t_m.

In the original BET method, a spherical brain was assumed and the brain volume was estimated taking account of all the voxels with intensities between t and t_{98} and the radius of the brain r was calculated. However, the length along each dimension of the brain is not always the same or even similar. For some datasets, there can be significant differences. It is more reasonable to assume an ellipsoidal brain. Thus in our work, we used half of r as the semi-major axis and half of the shortest edge of the bounding box as the semi-minor axis of the ellipsoid.

The brain surface model is initialised as a triangular tessellation of the spherical surface at first. It is centred at the COG and the radius is set to half of r. Thus the model is located inside of the brain initially. The surface is generated starting from an icosahedron. Then by iteratively subdividing each triangle into four smaller triangles and adjusting the positions of the vertices at the same time, the spherical surface is created. Given the semi-major and semi-minor axes calculated above, the coordinates of each vertex on the spherical surface are projected to the ellipsoidal surface, which is used as the initial model. The surface model initialisation process is illustrated in Fig. 4.2.

4.2.2 Surface Updating

In the original BET approach, the update equation used to perform the brain surface evolution is

$$\boldsymbol{u} = \boldsymbol{u_1} + \boldsymbol{u_2} + \boldsymbol{u_3}. \tag{4.2}$$

The role of $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$ is to maintain the smoothness of the evolving surface and keep the vertices equally distributed on the surface. They are derived from the difference vector between the current vertex and the mean position of its neighbours. $\boldsymbol{u_3}$ is the component which actually interacts with the brain image and forces the deforming model to get closer to and ultimately converge at the boundary of the brain surface. Fig. 4.3 illustrates $\boldsymbol{u_1}$, $\boldsymbol{u_2}$ and $\boldsymbol{u_3}$ calculated at one vertex on the deforming brain surface.

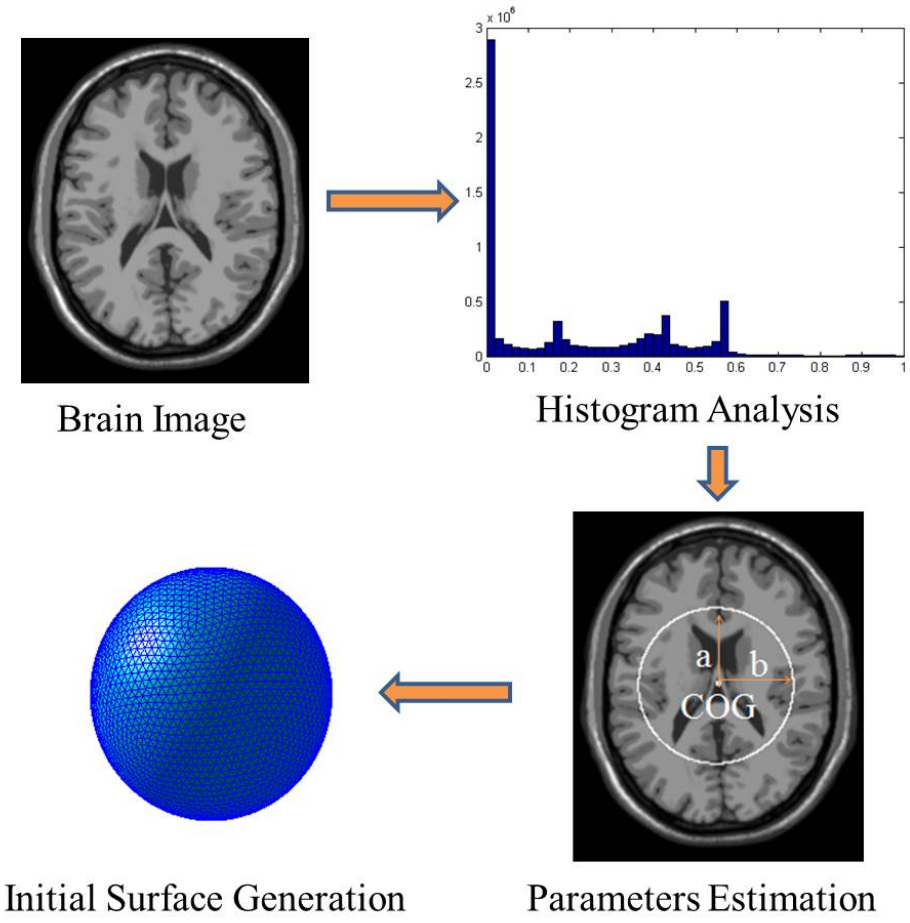


Fig. 4.2 Process of surface model initialisation. The parameters a and b represent the lengths of the semi-major and semi-minor axes, respectively. The original brain image is from BrainWeb.

$\boldsymbol{u_1}$ is tangential to the local surface; $\boldsymbol{u_2}$ and $\boldsymbol{u_3}$ are both along the normal direction of the local surface. In order to avoid the surface self-intersection, the displacements for the vertices at each iteration should be small enough. Thus the surface evolves slowly and the number of iterations required for a complete surface update is up to $1,000$. Even so, the deforming surface would lead to self-intersection and some regions have sharp boundaries that causes under-segmentation of the resulting brain as shown in Fig. 4.4. To avoid the weak points of the BET method, another way is proposed in our work for the surface updating. It performs surface deforming and reconstruction at each iteration. At the surface deforming step, only the force $\boldsymbol{u_3}$ is used and the vertex displacement is drastically increased. At the surface reconstruction step, instead of using $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$, we applied an independent surface reconstruction process to maintain the state of the deforming surface.

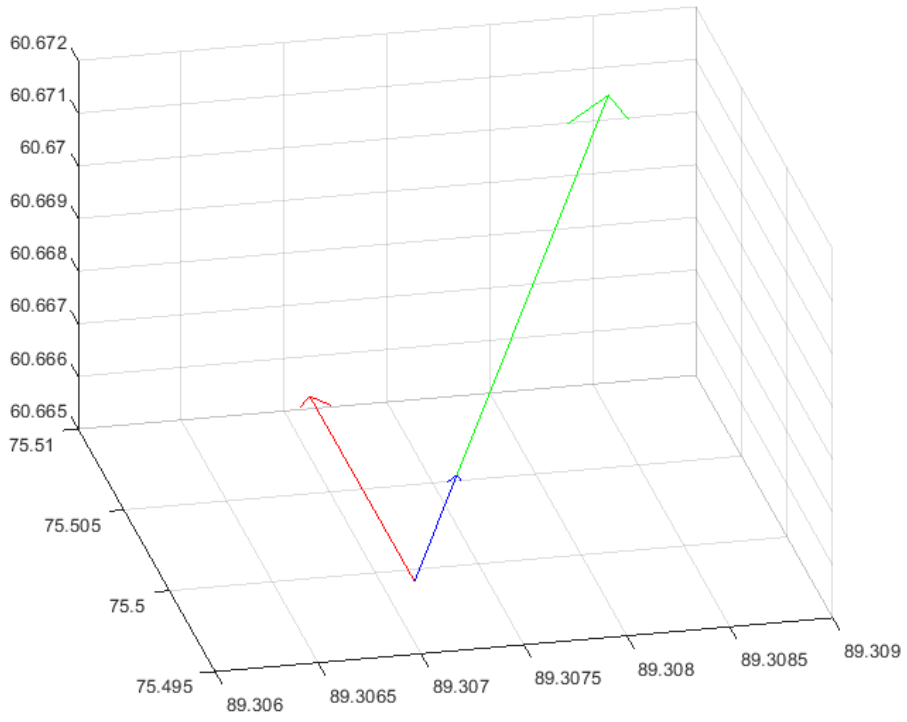


Fig. 4.3 Visualisation of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$ and $\boldsymbol{u_3}$ calculated at one vertex on the deforming brain surface generated by applying BET. $\boldsymbol{u_1}$, $\boldsymbol{u_2}$ and $\boldsymbol{u_3}$ are displayed in red, blue and green colours, respectively. Three axes represent the coordinates along three orthogonal directions.

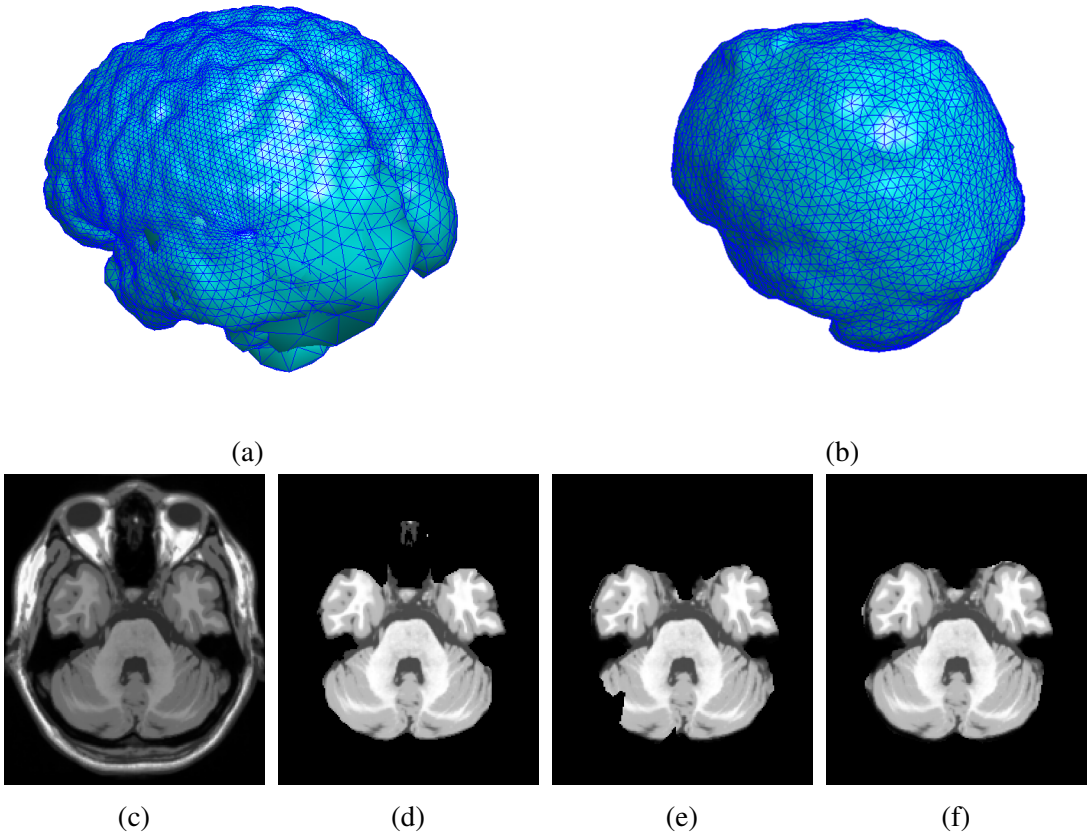
Surface deforming

Only $\boldsymbol{u_3}$ is used to drive the model to the brain surface. And the evolution speeds up by giving this term a scale coefficient λ which is set to 5 or 10 empirically in our experiments.Thus the update equation is

$$\boldsymbol{u} = \lambda \boldsymbol{u_3}. \tag{4.3}$$

The principle of $\boldsymbol{u_3}$ was originally proposed by Dale *et al.* [81] and later simplified by Smith in the BET method [263]. Considering the intensities of the voxels on the path from the current vertex pointing inward into the brain along the normal of the local surface (as path 1 shown in Fig. 4.5), the surface model is enforced to move inward or outward to reach the brain surface. In BET, $\boldsymbol{u_3}$ is given by

$$\boldsymbol{u_3} = 0.05 f_3 l \hat{\boldsymbol{s}}_{\boldsymbol{n}}. \tag{4.4}$$



(a) (b)

(c) (d) (e) (f)

Fig. 4.4 Comparison of brain surfaces generated from the original BET (a) and the improved BET (b) applying on a brain volume from BrainWeb. Brain images in the bottom row show an axial slice of the whole head image (c), the ground truth of brain segmentation (d), segmentation from the original BET (e) and segmentation from the improved BET (f).

The direction of $\boldsymbol{u_3}$ is parallel to $\hat{\boldsymbol{s}}_{\boldsymbol{n}}$-the unit normal at the current vertex. 0.05 is a weighting constant used in BET to provide a balance with $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$. l is the mean distance between the current vertex and its neighbours. f_3 is computed from the local intensity information

$$f_3 = \frac{2(I_{min} - t_1)}{I_{max} - t_2}, \tag{4.5}$$

where t_1 is the local threshold that distinguishes the brain from the background, which will be described later. The original BET computes local minimum and maximum intensities (I_{min} and I_{max}) as

$$\begin{cases} I_{min} = \max(t_2, \min(t_m, I(0), I(1), \ldots, I(d_1)) \\ I_{max} = \min(t_m, \max(t, I(0), I(1), \ldots, I(d_2)) \,, \end{cases} \tag{4.6}$$

where the intensity of the nth voxel on the search path from the current vertex is denoted as $I(n)$. d_1 and d_2 are the search lengths on the path to find I_{min} and I_{max}, d_2 is usually set to $d_1/2$. t_2, t_m and t are parameters obtained in surface model initialisation and here used to limit the effect of very dark or very bright voxels.

In this work, taking into account the complicated structure at the bottom of the brain, we add one more search path to calculate the local intensity parameters for the bottom slices (but no more than $1/3$ of the total number of slices) of the brain. The additional path also starts from the current vertex and points inward into the brain. However, its orientation is not along the normal but parallel to one of the x, y and z axes (as path 2 shown in Fig. 4.5). Given that the coordinates of the current vertex and COG in dimension i are $v(i)$ and $\mathrm{COG}(i)$ respectively, we compute

$$dim = \arg\max_i |v(i) - \mathrm{COG}(i)|, \quad i = 1,2,3. \tag{4.7}$$

Thus the orientation of the additional search path is parallel to the axis corresponding to the dimension dim. Then the new I_{min} and I_{max} are found:

$$\begin{cases} I_{min} = & \max(t_2, \min(t_m, I(0), I(1), \ldots, I(d_1), I'(0), I'(1), \ldots, I'(d_3)) \\ I_{max} = & \min(t_m, \max(t, I(0), I(1), \ldots, I(d_2), I'(0), I'(1), \ldots, I'(d_4))), \end{cases} \tag{4.8}$$

where $I'(n)$ is the intensity of the nth voxel from the current vertex on the additional path. And the search lengths for computing I_{min} and I_{max} on the additional path are set to d_3 and d_4, respectively. Similarly, d_4 is set to $d_3/2$ in the experiments.

Then the local intensity threshold t_1 is calculated as

$$t_1 = 0.5(I_{max} + t_2). \tag{4.9}$$

According to Eq. 4.5, the local minimum intensity I_{min} and local intensity threshold t_1 determine the sign symbol of f_3, which ranges roughly from -1 to 1. Hence, for Eq. 4.4, the direction of the force $\boldsymbol{u_3}$ is determined and the surface model is driven to move inward or outward to fit the brain surface eventually. Using one more searching path at the bottom of the brain, the intensity parameters are more robustly estimated and the deforming model evolves more correctly.

A comparison between the forces $\boldsymbol{u}$ calculated in the original BET and the improved BET is illustrated in Fig. 4.6.

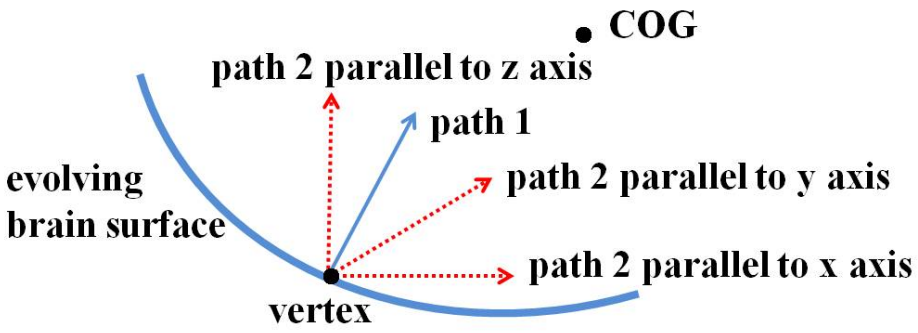
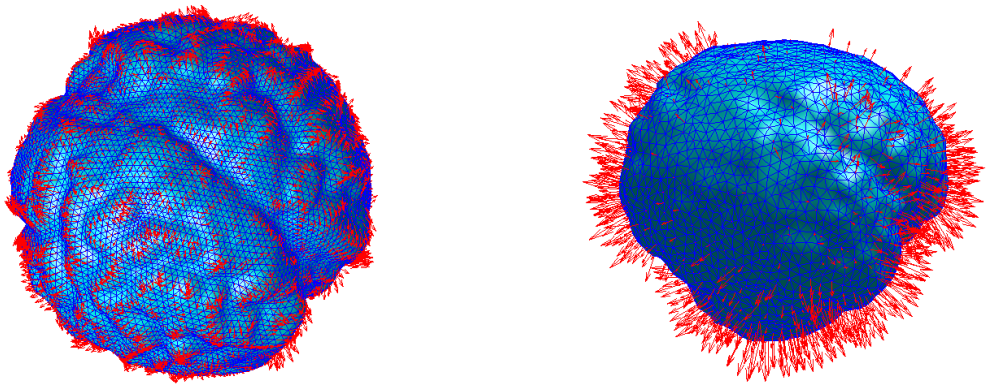


Fig. 4.5 The original and additional search paths denoted as path 1 and path 2, respectively.



(a) (b)

Fig. 4.6 Comparison of forces $\boldsymbol{u}$ calculated in the original BET (a) and the improved BET (b) applying on a brain volume from BrainWeb. For the purpose of visualisation, the magnitude of $\boldsymbol{u}$ for both methods is multiplied by 5.

Surface reconstruction

The deformed surface is reconstructed at each iteration and it includes two surface processes: self-intersection reduction and retriangulation.

To reduce surface self-intersection, we employ the method developed by Fang *et al.* [106]. First, the surface is converted to a volumetric image. Second, the holes of the volumetric image are filled. The filled volume is used to generate a new surface. The new vertices are generally well distributed over the new surface, where self-intersection is significantly reduced. However, the new surface produced by this method cannot guarantee consistent face orientation. We thus employ the algorithm proposed in [8] to retriangluate the vertices. Finally, we obtain a consistent surface without self-intersection and without holes.

In contrast to the final brain surface generated by the original BET method, the surface produced using this approach is not just of better quality, but more importantly, the number of iterations required for convergence is substantially decreased to less than 50. Fig. 4.4 shows a comparison of brain surfaces and the corresponding segmentation results generated by the original and the improved BET methods.

Examples of the brain evolving process of our method are illustrated in Figs 4.10(a) and 4.11.

Stopping criteria of surface updating

At each iteration, the volume enclosed in the mesh surface or the area of the mesh surface is calculated. When the difference of the mesh volume or the mesh surface area between the current and the previous iterations is small than a threshold ε_v or ε_s, the surface updating is considered converged.

4.2.3 False Positive Reduction

After brain segmentation, false positives are reduced using FCM [55] and some morphological operations. The extracted brain image is clustered into two categories by FCM according to the intensities of the voxels. Then the relatively dark voxels on the boundary of the brain are removed. By applying erosion, the remaining tissues are split into several separate regions and the largest connected region is selected as the brain. Finally, the selected brain region is dilated to compensate for the previous operations.

4.3 Experiments

We validated our method on both synthetic and real datasets. The results are compared with three other publicly available and commonly used brain segmentation methods. They are the original BET [263], BSE [252] and FreeSurfer's HWA [249]. The performance is evaluated both qualitatively and quantitatively. We will refer to the original BET method as BET and our improved BET method as BET_NEW in the following content.

4.3.1 Datasets

The synthetic dataset we used is BrainWeb MS lesion database [46]. We used 12 volumes with various levels of noise (0%, 3%, 5% and 7%) and intensity inhomo-

geneity (0%, 20% and 40%). The segments of WM, GM and CSF of the anatomical model are combined together to generate the ground truth. The second is IBSR18: 18 scans from healthy subjects are tested. These brain volumes are provided with manual annotations. The third dataset consists of 5 training volumes with manual labels from MICCAI2013 [195].

A detailed description of all the datasets was provided in Section 2.4. All the brain images tested in our experiments are T1-weighted as it provides high contrast between different tissues and the brain mask can be propagated to other modalities.

4.3.2 Evaluation Methods

Four metrics are computed: Sensitivity (SEN), Specificity (SPE), Dice similarity coefficient (DSC) [87] and misclassified tissue percentage (MTP). The SEN and SPE separately measure the proportion of brain and non-brain voxels correctly identified by the methods with respect to the ground truth. DSC measures the similarity between the segmentation result and the ground truth. These three metrics were defined in Table 3.1. MTP represents the percentage of potential brain voxels incorrectly included in or excluded from the segmentation result relative to the ground truth. It is defined as

$$MTP = \frac{FP+FN}{|A \bigcup B|} = \frac{FP+FN}{TP+FP+FN}, \tag{4.10}$$

where A and B are the sets of brain voxels in the segmentation result and the ground truth. For the metrics introduced above, the higher the value for the first three and the lower for the last one, the better the segmentation result.

In our experiments, the paired-sample t-test is adopted to analyse the impact of the parameter settings on the performance of our method. It measures the difference between two sets of segmentation results with the significance level set at 0.05. The repeated-measures MANOVA is used to compare the performance of different methods with respect to several metrics on the same database. Both of them are described in more detail in Section 3.5.

4.3.3 Parameters Optimisation

For BrainWeb and IBSR18, we performed k-fold cross validation to optimise the parameters for each method. The choice of k depends on the number of brain volumes in the database. We used 3-fold cross validation for BrainWeb (containing 12 volumes so 4 per fold) and IBSR18 (containing 18 volumes so 6 per fold). For MICCAI2013,

the leave-one-out cross validation was performed due to the low number of brain images.

For BET_NEW, three parameters were optimised: the scale coefficient λ and the length of two search paths d_1 and d_3. At first, we set initial values 20 and 0 for d_1 and d_3, respectively, as suggested in the original paper [263], and evaluated the performance of BET_NEW when varying λ. After applying k-fold cross validation or leave-one-out cross validation, the optimal λ was obtained. Then we set λ to the optimal value and tuned d_1 in a range around its initial value, for which we again used cross validation to find the optimal d_1. The parameter d_3 was optimised in a similar way. Finally, for each database, the brain segmentation results were produced from applying BET_NEW with the corresponding optimised parameters. We performed the same approach to optimise the parameter d_1 involved in BET.

Table 4.1 Optimised parameters from applying cross validation

Method		BET_NEW			BET
Parameter		λ	d_1	d_3	d_1
Database	BrainWeb	15, 20	24, 26, 30	6, 8	10
	IBSR18	5	20, 24	8, 10, 12	10
	MICCAI2013	5	12, 14	14	12

- λ represents the scale coefficient; d_1 and d_3 represent the length of the original and additional search paths, respectively.

The validation was repeated three times in optimising each parameter. It was found that the standard deviations among the segmentation results generated from repeating the optimisation process are very small.

In Table 4.1, we list the optimal values for all the parameters involved in both methods applied on various databases.

4.3.4 Validation of BET_NEW

In this section, we analyse the impact of main parameters involved in BET_NEW on the brain segmentation performance. The synthetic database BrainWeb is used for the validation. The algorithm reaches its convergence when the difference between the volumes enclosed in the mesh surfaces generated from the current and the previous iterations is less than 1000 (this value is obtained empirically).

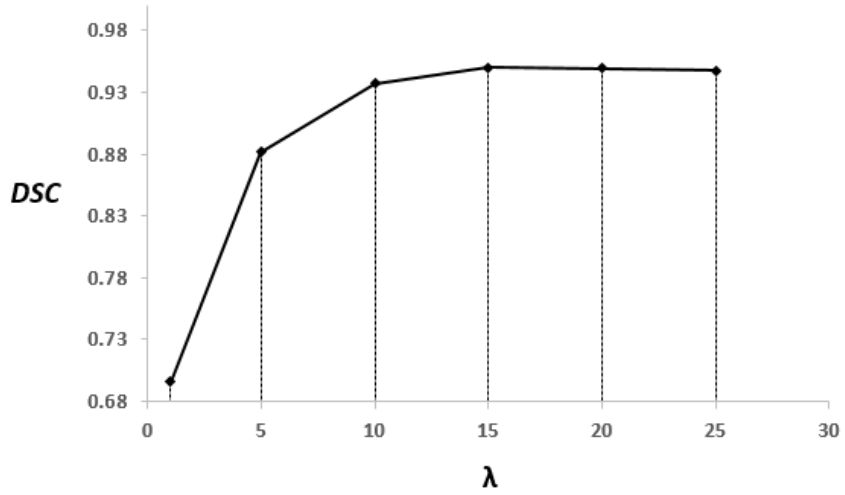


Fig. 4.7 Impact of λ on the performance of BET_NEW tested on BrainWeb and measured by DSC.

Impact of the scale coefficient λ

We set d_1 to 20 and d_3 to 0, respectively. The impact of λ on the segmentation performance of BET_NEW is analysed in Fig. 4.7 and Table 4.2. The segmentation accuracy measured by DSC is significantly improved when λ increases from 1 to 15. Then the DSC value does not change significantly when further increasing λ to 15 and 20 but slightly decreases when increasing λ to 25.

Impact of the search length d_1

We set the scale coefficient λ to its optimal value 15 and the length of the additional search path d_3 to 0. The length of the search path d_1 starts from 20 and varies in a range around it as shown in Fig. 4.8. Table 4.3 lists the segmentation accuracies of BET_NEW measured by DSC with different search lengths d_1. The segmentation accuracy is quite stable when gradually increasing d_1 from 20 to 38; it deteriorates when decreasing d_1 from 20 to 16.

Impact of the additional search path length d_3

Similarly, the impact of d_3 on the segmentation performance of BET_NEW is analysed in Fig. 4.9 and Table 4.3. We set λ to 15 and d_1 to 24. d_3 starts from 24, which is the optimal value for d_1, and gradually decreases. During this process, the segmentation accuracy measured by DSC is significantly improved until d_3 comes to 10. The DSC value is stable when further decreasing d_3 to 6. The best segmentation accuracy is

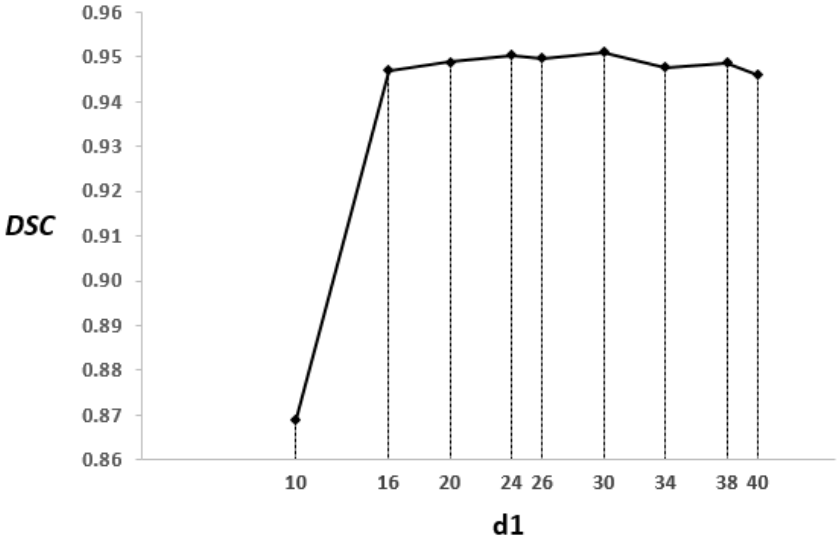


Fig. 4.8 Impact of d_1 on the performance of BET_NEW tested on BrainWeb and measured by DSC.

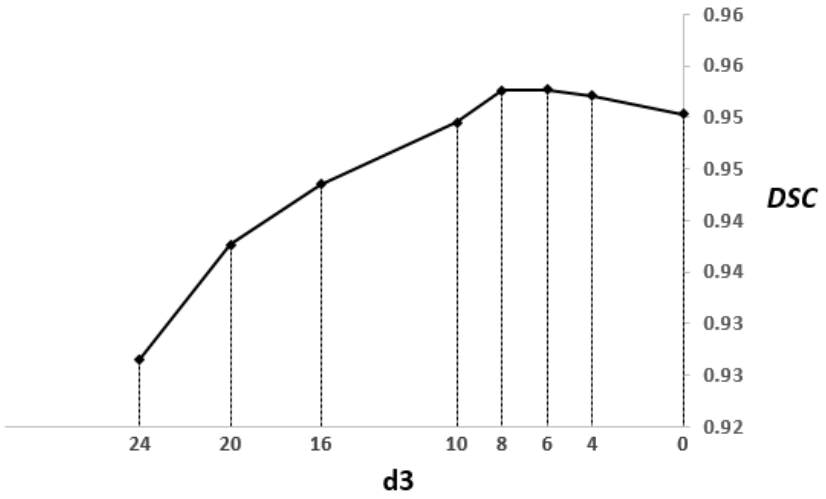


Fig. 4.9 Impact of d_3 on the performance of BET_NEW tested on BrainWeb and measured by DSC.

obtained when d_3 is 6 and this result is significantly better than that obtained without the additional search path (p-value < 0.05 as shown in Table 4.4). BET_NEW with the additional search path is applied only to the 50 (found empirically) bottom slices of the brain.

Parameter settings

We implemented both the original and the improved BET in MATLAB R2015a on a 64-bit Win7 operating system with Intel 3.00GHz CPU and 32.0GB RAM. For BET_NEW, different sets of optimal parameters are found by conducting parameter optimisation as demonstrated in Section 4.3.3. In the following experiments, the parameters of BET_NEW we used for each database are: $d_1 = 24$, $\lambda = 15$, $d_3 = 6$, $\varepsilon_v = 1000$ for BrainWeb; $d_1 = 20$, $\lambda = 5$, $d_3 = 8$, $\varepsilon_s = 1000$ for IBSR18; $d_1 = 12$, $\lambda = 5$, $d_3 = 14$, $\varepsilon_s = 100$ for MICCAI2013. BET_NEW with the additional search path is applied to only a certain number of bottom slices of the brain volumes in each database: 50 in BrainWeb, 60 in IBSR18 and 15 in MICCAI2013. These numbers together with the thresholds ε_v or ε_s are all empirically obtained. For BET, the values of d_1 we used for each database are 10, 10 and 12 for BrainWeb, IBSR18 and MICCAI2013, respectively. No suitable threshold ε_v or ε_s is found as the stopping criteria of the brain surface deforming and hence we simply set a threshold for the number of iterations: 1000 for BrainWeb; 600 for IBSR18; and 1000 for MICCAI2013. These numbers are all empirically obtained. For BSE, the BrainSuite13a1 software was downloaded and run on Windows. Four parameters including the diffusion iterations, the diffusion constant, the edge detection constant and the erosion size were fine tuned to generate acceptable result for each image. The options of 'trim brain stem' and 'dilate final mask' were set as 'true' for all the images. For HWA, the Linux version of FreeSurfer v5.3.0 was run on a virtual box. The command 'mri_watershed' was applied to perform the segmentation for all the images.

Table 4.2 Impact of λ on the performance of BET_NEW tested on BrainWeb and measured by DSC

	1	5	10	15	20	25
1	$\mathbf{0.6960 \pm 0.0107}$	$3.02e-14^{\star}$	$4.46e-15^{\star}$	$2.56e-16^{\star}$	$9.53e-17^{\star}$	$8.75e-17^{\star}$
5	–	$\mathbf{0.8819 \pm 0.0060}$	$2.60e-14^{\star}$	$1.77e-12^{\star}$	$4.08e-11^{\star}$	$9.60e-11^{\star}$
10	–	–	$\mathbf{0.9374 \pm 0.0071}$	$4.94e-05^{\star}$	$0.0012^{\star}$	$0.0045^{\star}$
15	–	–	–	$\mathbf{0.9498 \pm 0.0030}$	0.6256	0.0548
20	–	–	–	–	$\mathbf{0.9493 \pm 0.0034}$	$0.0007^{\star}$
25	–	–	–	–	–	$\mathbf{0.9474 \pm 0.0042}$

- The numbers in the first column and the first row are the λ values tested for BET_NEW applied on BrainWeb. The values highlighted in bold show the mean DSC values and the standard deviations produced by BET_NEW with the corresponding λ. Each of the other numbers in the table shows the p-value tested between two sets of segmentation results produced with a pair of λ values in the corresponding row and column. The significant differences are indicated by $\star$ at the top right of the p-values.
- The p-values in the table are symmetric, so we only show half of them and the omitted ones are replaced by '–'.

Table 4.3 Impact of d_1 on the performance of BET_NEW tested on BrainWeb and measured by DSC

	10	16	20	24	26	30	38
10	$\mathbf{0.8690 \pm 0.0853}$	$0.0068^{\star}$	$0.0062^{\star}$	$0.0058^{\star}$	$0.0080^{\star}$	$0.0066^{\star}$	$0.0092^{\star}$
16	–	$\mathbf{0.9468 \pm 0.0062}$	0.0531	$0.0164^{\star}$	0.3504	0.0509	0.5384
20	–	–	$\mathbf{0.9488 \pm 0.0045}$	$0.0035^{\star}$	0.7115	0.1464	0.9551
24	–	–	–	$\mathbf{0.9503 \pm 0.0037}$	0.8021	0.5420	0.4143
26	–	–	–	–	$\mathbf{0.9497 \pm 0.0069}$	0.5192	0.6535
30	–	–	–	–	–	$\mathbf{0.9511 \pm 0.0014}$	0.1072
38	–	–	–	–	–	–	$\mathbf{0.9486 \pm 0.0048}$

- The numbers in the first column and the first row are the d_1 values tested for BET_NEW applied on BrainWeb. The values highlighted in bold show the mean DSC values and the standard deviations produced by BET_NEW with the corresponding d_1. Each of the other numbers in the table shows the p-value tested between two sets of segmentation results produced with a pair of d_1 values in the corresponding row and column. The significant differences are indicated by $\star$ at the top right of the p-values.
- The p-values in the table are symmetric, so we only show half of them and the omitted ones are replaced by '–'.

Table 4.4 Impact of d_3 on the performance of BET_NEW tested on BrainWeb and measured by DSC

	24	20	16	10	8	6	0
24	$\mathbf{0.9265 \pm 0.0160}$	$0.0092^{\star}$	$0.0009^{\star}$	$5.07e-05^{\star}$	$0.0002^{\star}$	$0.0002^{\star}$	$0.0009^{\star}$
20	–	$\mathbf{0.9377 \pm 0.0143}$	$0.0342^{\star}$	$0.0069^{\star}$	$0.0032^{\star}$	$0.0032^{\star}$	$0.0220^{\star}$
16	–	–	$\mathbf{0.9435 \pm 0.0129}$	0.1226	$0.0280^{\star}$	$0.0268^{\star}$	0.1366
10	–	–	–	$\mathbf{0.9495 \pm 0.0070}$	0.1731	0.1637	0.7697
8	–	–	–	–	$\mathbf{0.9526 \pm 0.0014}$	0.8825	0.0657
6	–	–	–	–	–	$\mathbf{0.9527 \pm 0.0013}$	$0.0444^{\star}$
0	–	–	–	–	–	–	$\mathbf{0.9503 \pm 0.0037}$

- The numbers in the first column and the first row are the d_3 values tested for BET_NEW applied on BrainWeb. The values highlighted in bold show the mean DSC values and the standard deviations produced by BET_NEW with the corresponding d_3. Each of the other numbers in the table shows the p-value tested between two sets of segmentation results produced with a pair of d_3 values in the corresponding row and column. The significant differences are indicated by $\star$ at the top right of the p-values.
- The p-values in the table are symmetric, so we only show half of them and the omitted ones are replaced by '–'.

4.3.5 Segmentation Performance Comparisons

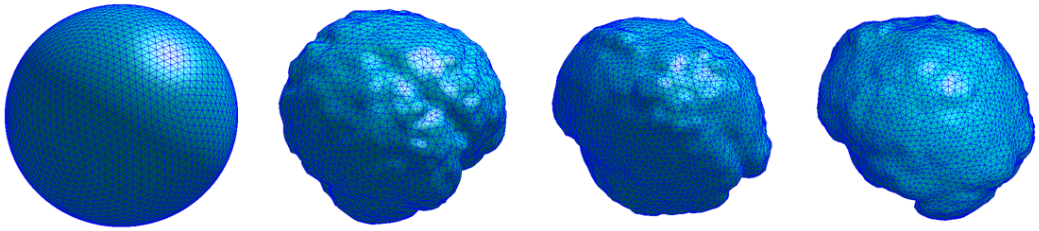
Comparisons between BET and BET_NEW

The brain surface deforming processes of applying BET and BET_NEW to a brain volume from BrainWeb are compared in Fig. 4.10. Owing to the surface reconstruction step embedded in each iteration of BET_NEW, the evolving brain surfaces are more smooth and the vertices are more evenly distributed on the surfaces. In contrast, the brain surfaces produced by BET show more details, which results in better specificity in brain segmentation. However, the overall segmentation performance is deteriorated from surface self-intersection and poor distribution of the vertices. The brain surface deforming processes of applying BET_NEW on brain volumes from IBSR18 and MICCAI2013 are also illustrated in Fig. 4.11.

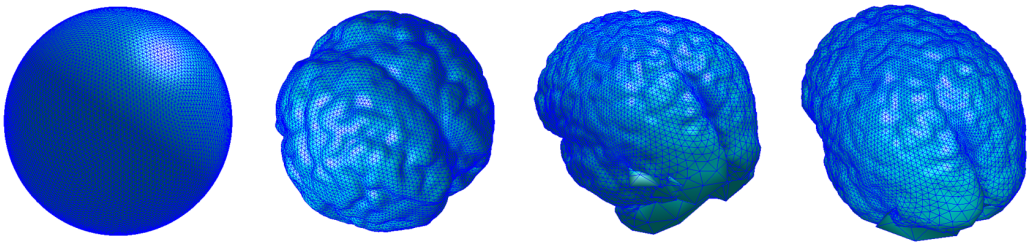
A qualitative comparison is shown in Fig. 4.12. The accuracies of each segmented brain volume shown in Fig. 4.12 measured by four metrics are listed in Table 4.6. Comparing the segmentation results of BET (Figs 4.12(g-i)) with the ground truth (Figs 4.12(d-f)), BET tends to undersegment the brain volumes from BrainWeb and MICCAI2013. In the meanwhile, it oversegments the brain volumes from IBSR18. From Figs 4.12(j-l), we can see that BET_NEW oversegments the brain volumes from all three databases; the oversegmentation of BET_NEW is slightly less than BET for IBSR18. The overall segmentation performance of BET_NEW is better than that of BET, which is reflected by the segmentation accuracies measured by DSC and MTP shown in Table 4.7.

The brain segmentation accuracies of BET and BET_NEW are quantitatively compared in Tables 4.7 and 4.8. The mean accuracies and the standard deviations of applying both methods on the whole database measured by each metric are listed in Tables 4.7. In Table 4.8, the paired-sample t-test is applied to test the difference between the segmentation performance of BET and BET_NEW with respect to each metric. We found the quantitative comparison results are quite consistent with the qualitative comparisons: in the segmentation of BrainWeb and MICCAI2013, BET_NEW obtains higher sensitivity and BET obtains higher specificity; in the segmentation of IBSR18, BET_NEW shows higher specificity. Taking all metrics into account, BET_NEW performs significantly better than BET with respect to three metrics (SEN, DSC and MTP) for BrainWeb and MICCAI2013; it also significantly outperforms BET in terms of one metric (SPE) for IBSR18.

We also compared the number of iterations and the total execution time required for both methods applying on each database in Table 4.5. As described in Section 'Parameter settings', the stopping criteria of brain surface deforming using BET_NEW

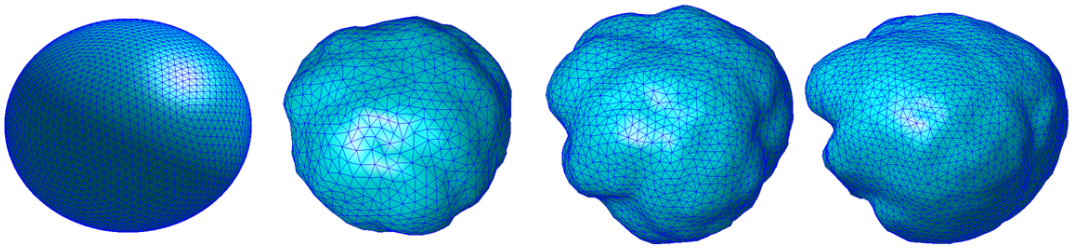


(a) Brain surface evolution at iterations 0, 10, 20 and 33.

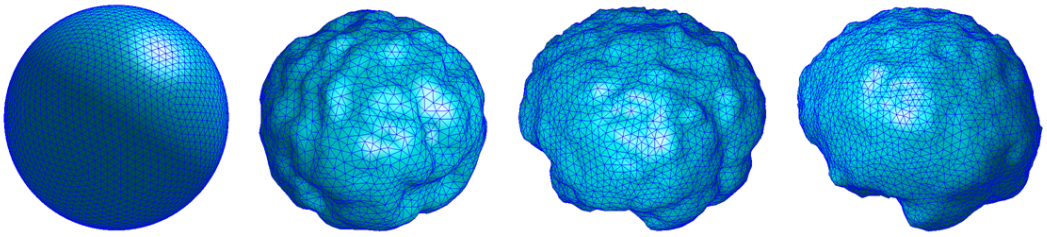


(b) Brain surface evolution at iterations 0, 100, 600 and 1,000.

Fig. 4.10 Comparison between evolving brain surfaces produced by BET_NEW(a) and BET(b) applying to a volume from BrainWeb.



(a) Brain surface evolution at iterations 0, 10, 20 and 24.



(b) Brain surface evolution at iterations 0, 10, 20 and 34.

Fig. 4.11 Deforming brain surfaces produced by BET_NEW applying to brain volumes from IBSR18(a) and MICCAI2013(b).

Table 4.5 Comparisons of the algorithm efficiency between BET and BET_NEW

Database	Method	Number of iterations	Execution time (s)
BrainWeb	BET	1,000	3936.8645 ± 70.2618
	BET_NEW	45 ± 6	391.928 ± 65.1058
IBSR18	BET	600	2044.0865 ± 2.0981
	BET_NEW	26 ± 3	112.7934 ± 29.7313
MICCAI2013	BET	1,000	3014.626 ± 57.7294
	BET_NEW	35 ± 2	242.548 ± 21.2333

- Each value represents the average number of iterations or execution time (with the standard deviation) required for each method applied on each database.

is determined by a threshold calculated from the mesh volume difference ε_v or the mesh surface area difference ε_s between two successive iterations. For BET, we empirically set the numbers of iterations required for the three databases. From Table 4.5, we can see that due to the enhancement of vertex displacement during the iterative surface evolution, the number of iterations needed is dramatically reduced. Moreover, the execution time is also substantially decreased by more than 10 times.

Comparisons with other methods

In addition to BET, we also compared the segmentation results produced by BET_NEW with those produced by BSE and HWA. The example slices of the segmented brain volumes by applying each method on different databases are shown in Fig. 4.12. The corresponding accuracies of each segmented brain volume measured by four metrics are listed in Table 4.6. It is observed that BSE undersegments the brain images from all three databases, which is also indicated by its lowest sensitivity in the segmentation of BrainWeb and MICCAI2013 and highest specificity obtained for all three databases. For the brain volumes shown in Fig. 4.12, BSE produces the best results for BrainWeb and IBSR18 in terms of DSC and MTP. However, the undersegmentation is so severe that it produces the worst segmentation for MICCAI2013. On the contrary, HWA oversegments the brain volumes from all three databases, which is also indicated by its almost highest sensitivity (for MICCAI2013, it produces the second highest) and lowest specificity in the segmentation of each database. Considering DSC and MTP, its segmentation performance is the worst for BrainWeb and IBSR18 and the second worst for MICCAI2013. BET and BET_NEW generate relatively acceptable segmentation results for all three databases.

A quantitative comparison of the brain segmentation performance of four different methods on three databases was conducted and the results are listed in Table 4.7. The repeated-measures MANOVA was applied to test the differences between the four methods. The multivariate outcomes indicate that significant differences exist between the four methods applied to each database in terms of the combination of the four metrics. The univariate outcomes demonstrate that significant difference exists among the four methods in terms of each metric in the segmentation of brain volumes from each database. The *post hoc* pairwise comparisons of the four methods were conducted with respect to each metric on each database. From Table 4.7, conclusions consistent with the qualitative comparison can be drawn: HWA produces the highest sensitivity and the lowest specificity for the three databases; conversely, BSE obtains the lowest sensitivity for two databases (except for IBSR18) and the highest specificity for three databases. In terms of DSC and MTP, HWA performs the worst in the segmentation of BrainWeb and IBSR18; BSE performs the best in the segmentation of IBSR but the worst for MICCAI2013; BET_NEW performs the best for BrainWeb and MICCAI2013. Taking all the four metrics into account, we can draw the conclusions that for BrainWeb, HWA produces the worst segmentation; for IBSR18, BSE performs the best; for MICCAI2013, BSE obtains the worst segmentation; BET generates moderate segmentation results for all three databases.

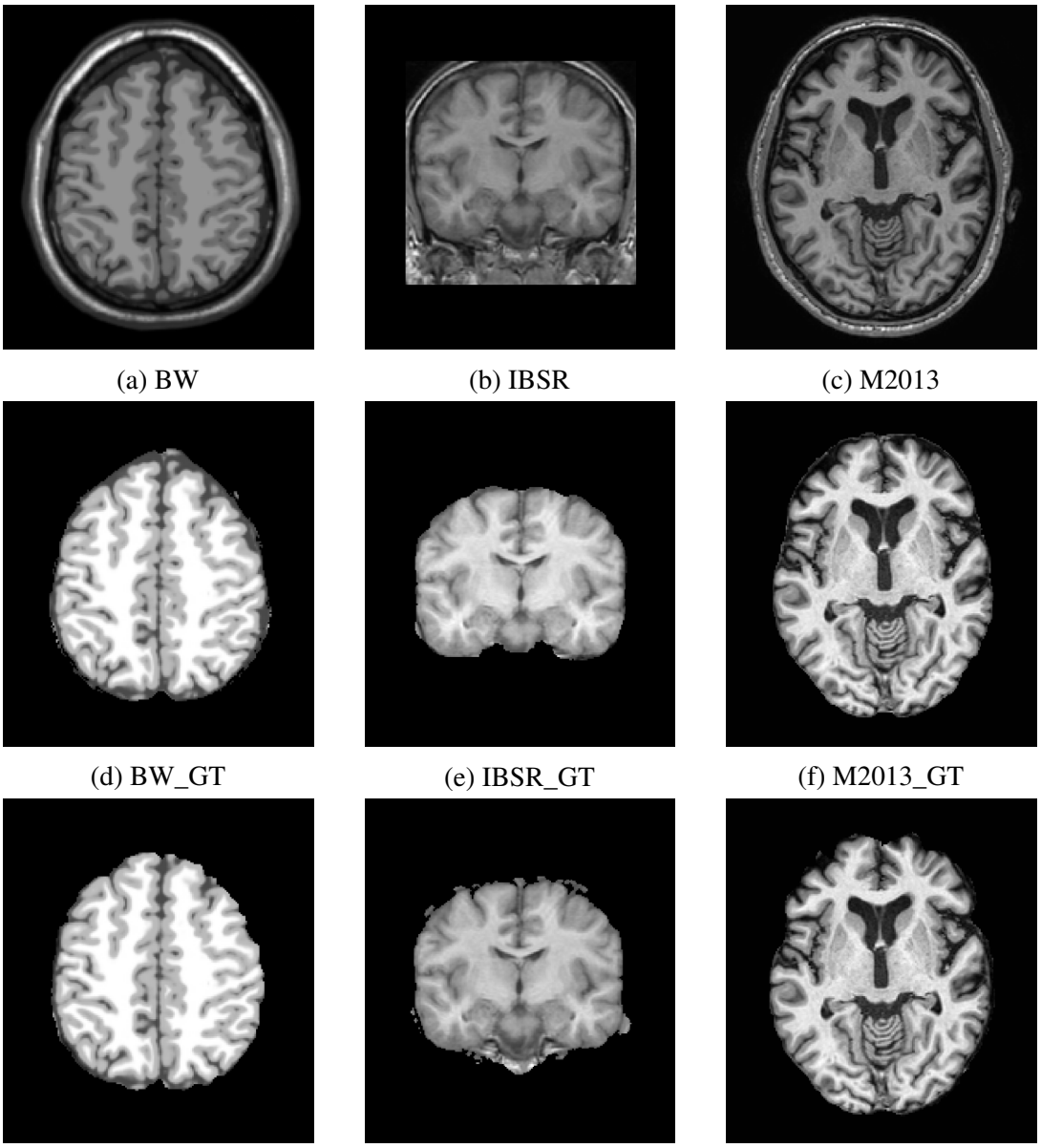
In order to compare our method with each of the comparison methods, the paired-sample t-test is used to test the difference between BET_NEW and each of the other methods with respect to each metric for each database. From the comparison results shown in Table 4.8, we can see that BET_NEW significantly outperforms BET in terms of at least two metrics in the segmentation of three databases. By contrast with BSE, BET_NEW produces comparable segmentation results for BrainWeb; it performs significantly worse in terms of DSC and MTP for IBSR18; however, it obtains significantly better segmentation accuracy in terms of SEN, DSC and MTP for MICCAI2013. Compared to HWA, BET_NEW produces significantly better results with respect to three metrics (SEN, DSC and MTP) in the segmentation of three databases. Comparing the performance between BET_NEW and the other three methods in the segmentation of each database and taking all the four metrics into account, we can also conclude that BET_NEW outperforms the other three methods for BrainWeb and MICCAI2013; it performs the second best for IBSR18; BSE generates the best segmentation results for IBSR18.

4.3.6 Discussion

The original BET method produces acceptable segmentation results for brain volumes from all three databases. However, it requires a large number of iterations for the deforming brain surface to reach convergence, which directly results in a long processing time needed to segment a brain volume. It tends to undersegment the brain volumes from BrainWeb and MICCAI2013 and oversegments the brain volumes from IBSR18. The surfaces generated from BET show more details of the brain surfaces. However, because it does not efficiently deal with the mesh self-intersection, the produced brain surfaces are not smooth enough and the vertices are not evenly distributed over the surfaces, which also deteriorates the segmentation performance. BSE undersegments the brain volumes from all three databases. It generates the best segmentation results for IBSR18 but the worst for MICCAI2013. Its performance highly relies on the parameter settings: in our experiments, we fine tuned the parameters for each brain volume to produce acceptable segmentation. Contrary to BET and BSE, HWA oversegments the brain volumes from all three databases, which makes it perform the worst in the segmentation of BrainWeb and IBSR18.

BET_NEW significantly outperforms BET in the segmentation of brain volumes from all three databases. The efficiency and the improved accuracy of BET_NEW owe more to the surface reconstruction process and the FCM based false positive reduction. The enhancement of the vertex displacement at each iteration largely decreases the total number of iterations required for a complete brain surface evolution. The employment of an independent surface reconstruction step solves the problem of mesh self-intersection and generates smooth brain surfaces at the same time. Furthermore, the execution time is substantially reduced. The additional search path makes the estimation of the local intensity threshold at the bottom of the brain more accurate. Compared to BSE, it produces better segmentation results in the segmentation of brain volumes from BrainWeb and MICCAI2013 but it performs worse for IBSR18. In contrast to HWA, BET_NEW produces significantly better segmentation results for all three databases.

BET_NEW works well in the middle or at the top of the brain volumes from different databases. But at the bottom of the brain where the structure is more complicated and the brain boundaries are ambiguous, it could fail to completely remove the eyeballs and produce oversegmentation as illustrated in Fig. 4.13. On the other hand, the surface reconstruction step is a little time consuming and it takes most of the time in the whole process of brain segmentation.

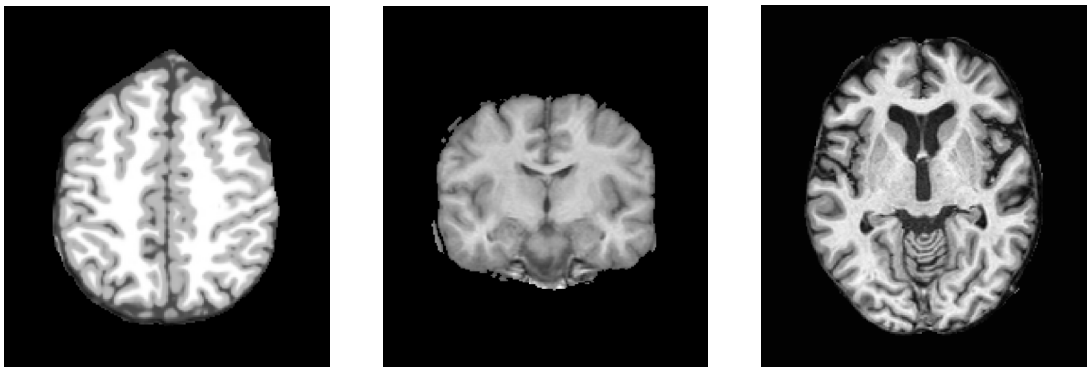


(a) BW (b) IBSR (c) M2013

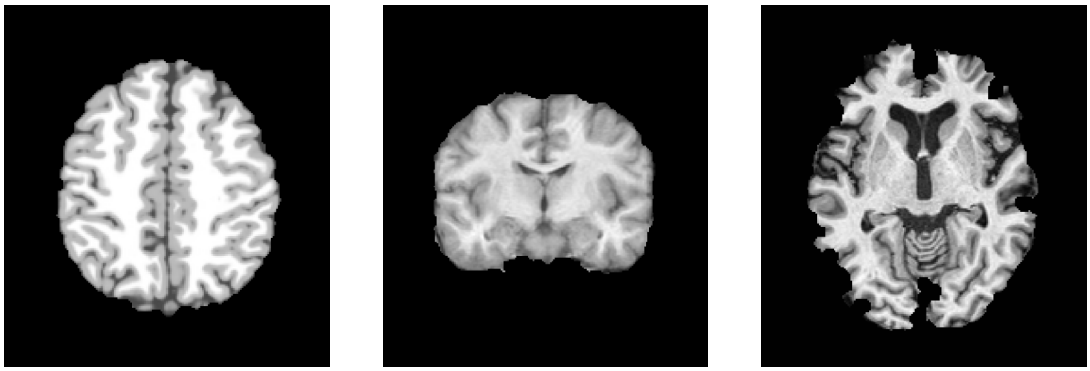
(d) BW_GT (e) IBSR_GT (f) M2013_GT

(g) BW_BET (h) IBSR_BET (i) M2013_BET

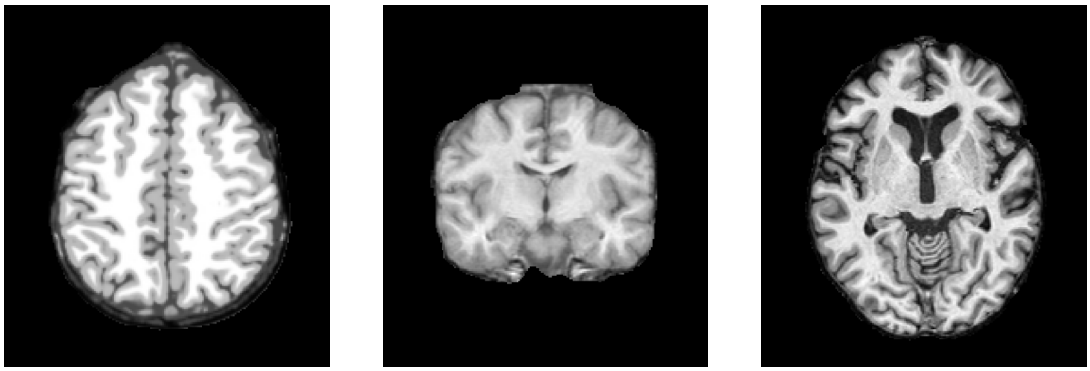
Fig. 4.12 Qualitative comparisons of brain segmentation using 4 different methods. Images (a-c) are the original slices of brain volumes from BrainWeb (BW), IBSR18 (IBSR) and MICCAI2013 (M2013); their corresponding ground truth (GT) of brain segmentation are (d-f); images (g-i) are the segmentation results by applying BET. The segmentation accuracies of the brain volumes measured by SEN, SPE, DSC and MTP are listed in Table 4.4.



(j) BW_BET_NEW (k) IBSR_BET_NEW (l) M2013_BET_NEW



(m) BW_BSE (n) IBSR_BSE (o) M2013_BSE



(p) BW_HWA (q) IBSR_HWA (r) M2013_HWA

Fig. 4.12 Qualitative comparisons of brain segmentation using 4 different methods continued. Images (j-l) represent the segmentation results by BET_NEW; (m-o) are the results from applying BSE; (p-r) are produced by HWA. The segmentation accuracies of the brain volumes measured by SEN, SPE, DSC and MTP are listed in Table 4.4.

Table 4.6 Segmentation accuracies of the brain volumes shown in Fig. 4.12

Brain volume	*SEN*	*SPE*	*DSC*	*MTP*
BW_BET(g)	0.9336	0.9846	0.9460	0.1067
BW_BET_NEW(j)	0.9825	0.9688	0.9520	0.0991
BW_BSE(m)	0.9286	0.9917	0.9522	0.0932
BW_HWA(p)	0.9950	0.9356	0.9197	0.1737
IBSR_BET(h)	0.9598	0.9915	0.9534	0.0938
IBSR_BET_NEW(k)	0.9558	0.9938	0.9584	0.0830
IBSR_BSE(n)	0.9764	0.9950	0.9725	0.0553
IBSR_HWA(q)	0.9977	0.9735	0.9218	0.1692
M2013_BET(i)	0.9445	0.9825	0.9337	0.1342
M2013_BET_NEW(l)	0.9883	0.9741	0.9391	0.1282
M2013_BSE(o)	0.8406	0.9961	0.9047	0.1771
M2013_HWA(r)	0.9828	0.9709	0.9301	0.1478

- Numbers in the table represent the segmentation accuracies of each brain volume shown in Fig. 4.12 measured by SEN, SPE, DSC and MTP, respectively.

Table 4.7 Comparisons of the performance of different brain segmentation methods

Database	Method	*SEN*	*SPE*	*DSC*	*MTP*
BrainWeb	BET	$0.9543 \pm 0.0254^{\star}$	$0.9763 \pm 0.0088^{\star}$	$0.9465 \pm 0.0086^{\star}$	$0.1077 \pm 0.0162^{\star}$
	BET_NEW	$0.9609 \pm 0.0147^{\star}$	$0.9779 \pm 0.0064^{\star}$	$0.9520 \pm 0.0014^{\star}$	$0.0970 \pm 0.0034^{\star}$
	BSE	0.9115 ± 0.0165	$0.9946 \pm 0.0019^{\star\star}$	$0.9466 \pm 0.0076^{\star}$	$0.1026 \pm 0.0135^{\star}$
	HWA	$0.9922 \pm 0.0022^{\star\star}$	0.9378 ± 0.0018	0.9209 ± 0.0012	0.1706 ± 0.0031
IBSR18	BET	0.9378 ± 0.0387	$0.9841 \pm 0.0083^{\star}$	0.9213 ± 0.0243	0.1601 ± 0.0479
	BET_NEW	0.9535 ± 0.0120	$0.9978 \pm 0.0042^{\star\star\star}$	$0.9386 \pm 0.0107^{\star}$	$0.1249 \pm 0.0228^{\star}$
	BSE	0.9576 ± 0.0235	$0.9939 \pm 0.0050^{\star\star}$	$0.9583 \pm 0.0154^{\star\star}$	$0.0838 \pm 0.0330^{\star\star}$
	HWA	$0.9909 \pm 0.0056^{\star}$	0.9744 ± 0.0110	0.9225 ± 0.0182	0.1675 ± 0.0438
MICCAI2013	BET	0.9177 ± 0.0208	$0.9888 \pm 0.0037^{\star}$	$0.9328 \pm 0.0064^{\star}$	$0.1321 \pm 0.0116^{\star}$
	BET_NEW	$0.9784 \pm 0.0089^{\star}$	0.9785 ± 0.0037	$0.9439 \pm 0.0030^{\star}$	$0.1163 \pm 0.0071^{\star}$
	BSE	0.8386 ± 0.0281	$0.9940 \pm 0.0026^{\star}$	0.8994 ± 0.0142	0.1871 ± 0.0232
	HWA	$0.9802 \pm 0.0072^{\star}$	0.9716 ± 0.0062	$0.9315 \pm 0.0057^{\star}$	$0.1443 \pm 0.0138^{\star}$

- Each value represents the mean accuracy and the standard deviation of applying each method on the database.
- The $\star$s at the top right of each value indicate the significant differences of the corresponding result compared with all the others derived by applying other methods on the same database and measured by the same metric. The result labelled with more $\star$s is significantly better than those with fewer or no $\star$s and the results labelled with equal number of $\star$s have no significant difference between each other.

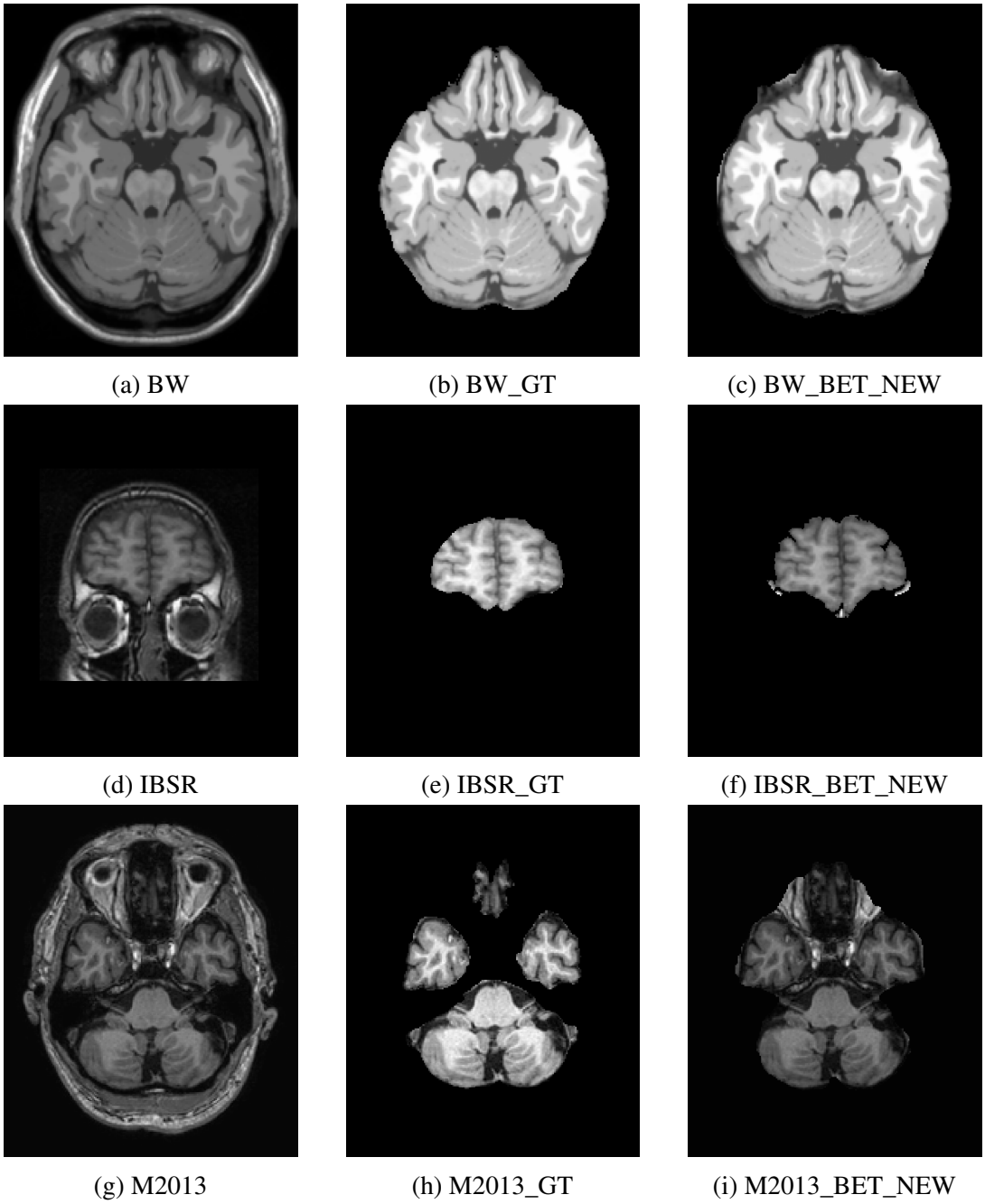
Table 4.8 Comparisons of the segmentation performance between BET_NEW and other methods

Database	Method	*SEN*	*SPE*	*DSC*	*MTP*
BrainWeb	BET	0.2520	0.1295	0.0423^{+}	0.0305^{+}
	BSE	$3.18e-06^{+}$	$1.16e-07^{-}$	0.0556	0.2402
	HWA	$3.49e-06^{-}$	$1.38e-11^{+}$	$1.90e-14^{+}$	$4.82e-15^{+}$
IBSR18	BET	0.0466^{+}	0.0058^{+}	$1.73e-04^{+}$	$9.08e-05^{+}$
	BSE	0.5382	$4.03e-04^{+}$	$3.88e-04^{-}$	$4.08e-04^{-}$
	HWA	$2.27e-10^{-}$	$7.00e-06^{+}$	0.0019^{+}	$6.94e-04^{+}$
MICCAI2013	BET	0.0038^{+}	0.0034^{-}	0.0215^{+}	0.0382^{+}
	BSE	$2.07e-04^{+}$	$6.40e-04^{-}$	0.0021^{+}	0.0027^{+}
	HWA	0.5380	$3.46e-02^{+}$	0.0093^{+}	0.0113^{+}

- Each value represents the p-value calculated from comparing the segmentation accuracy between BET_NEW and each of other methods applying on the same database measured by each metric.
- The symbol '+' or '-' at the top right of the p-value indicates BET_NEW significantly outperforms or underperforms the method compared with in terms of the corresponding metric; the p-value labelled with no '+' or '-' indicates no significant difference exists between BET_NEW and the method compared with in terms of the corresponding metric.

4.4 Conclusions

We have presented an improved BET method for brain segmentation. Compared to the original BET method, we added one more search path at the bottom of the brain to make the segmentation more robust. At each iteration, the vertex displacement was enhanced and we employed an independent surface reconstruction process to maintain the smoothness of the surface and the equal distribution of the vertices on the surface. At the end, a FCM based scheme was used to reduce the false positives. Our approach dramatically decreased the number of iterations for surface updating. The qualitative and quantitative comparisons were conducted with three commonly used brain segmentation methods on three publicly available datasets. It has been shown that most of the segmentation results generated by our method are better than or comparable to those produced by the other methods. Future work will focus on improving the segmentation at the bottom of the brain volume and simplifying the process of surface reconstruction.



(a) BW (b) BW_GT (c) BW_BET_NEW

(d) IBSR (e) IBSR_GT (f) IBSR_BET_NEW

(g) M2013 (h) M2013_GT (i) M2013_BET_NEW

Fig. 4.13 Examples of bottom slices of brain volumes segmented by BET_NEW. Images (a, d, g) are original slices of brain volumes from BrainWeb(BW), IBSR18(IBSR) and MICCAI2013(M2013); their corresponding ground truth of brain segmentation are shown in (b, e, h); images (c, f, i) are segmentation results by applying BET_NEW.

Chapter 5

Brain Image Registration

5.1 Introduction

Brain image registration aims to find a transformation that brings the moving image (the image to register) into a voxel-to-voxel correspondence with the static image (the target image). According to the nature of the transform, registration methods can be categorised into rigid/affine and non-rigid approaches. Rigid and affine registrations are usually applied to pre-register the brain images. And the more precise matching between internal structures is achieved by non-rigid registration. In terms of the features used for registration, these methods are summarised into landmark based, segmentation based, intensity based and a hybrid of two or three of the methods above [116]. Salient points like edges are detected as landmarks to perform registration [236]. Segmentation based methods use internal structures, curves or surfaces for registration [82]. Intensity is the most important feature in brain image registration [152, 233]. In recent years, a hybrid of intensity information, landmarks and surfaces has been widely used, showing promising results [211, 142].

5.1.1 Previous Work

Considering the complex internal structures in brain images, non-rigid registration has to be performed to match two brains. Numerous brain image registration techniques have been proposed in the literature, which vary in features used, transformation model, similarity measure, optimisation method and resampling scheme [284, 77, 151].

As reviewed in Section 3.3.2, the transformation models used in image registration include the polynomial transformations, radial basis functions and elastic models. For brain image registration, an automated image registration (AIR) method [316] was

proposed to match two images based on voxel intensities by minimising the mean squared difference between them. The polynomial transformation is estimated with its order implemented up to twelve. Registration starts from the first order polynomial and proceeds to the next higher order polynomials when it satisfies the criteria such as the improvement reflected in the cost function or the number of iterations. A consistent linear-elastic brain image registration method was proposed in [63]. It jointly estimates a consistent set of forward and reverse transformations between two images by minimising a cost function, which is defined by the intensity correspondence, the displacement field regularisation term and the inverse consistency constraint. The transformation is parameterised based on a 3D Fourier series in the context of linear elasticity and optimised using the gradient descent method. A nonlinear spatial normalisation method embedded in the statistical parametric mapping (SPM5) package was developed by Ashburner et al. [18]. An affine registration is firstly applied with a Bayesian framework to estimate 12 parameters. Subsequently a nonlinear registration is used, whereby the transformation is represented by a linear combination of 3D discrete cosine transform basis functions. The coefficients for each basis function are optimised by simultaneously minimising the sum of squared differences between two images and maximising the smoothness of the transformation field. A non-rigid registration using free-from deformations was developed to match two 3D breast MR images [239]. A hierarchical transformation model is used, where the global transformation is modelled by an affine transformation while the local transformation is modelled by a free-form deformation based on B-splines. The transformation is optimised by a gradient descent minimising a cost function, which combines the cost associated with the image similarity measured by normalised mutual information and the cost associated with the smoothness of the transformation. A nonlinear brain registration algorithm included in the automated registration toolbox [13] estimates a non-parametric free-form transformation between the moving and static images using the local normalised cross-correlation as its similarity measure. A 6-parameter rigid registration followed by an affine registration is optional for the toolbox. It was claimed by the authors that this algorithm outperforms both AIR and the nonlinear spatial normalisation method in SPM5 in reducing inter-subject anatomical variability. A toolbox named elastix [152] for intensity-based medical image registration was developed to provide more convenience for configuring, testing and comparing different registration methods. It consists of a collection of algorithms with different choices for each component of registration: the cost function, the transformation model, the optimisation method, the image sampling strategy, the interpolation method and the hierarchical strategy. A fluid warping technique was proposed in [62] to perform tensor-based

morphometry. Images are fluidly registered by applying the driving forces throughout the moving image to maximise the Jensen-Rényi divergence between the two images.

Various features have been used to perform brain image registration. A fully automated spatial normalisation approach, referred to as hierarchical attribute matching mechanism for elastic registration (HAMMER), was proposed for volumetric morphometry [255]. It uses a collection of geometric attributes and a hierarchical approximation of the similarity function in the registration process. A mass-preserving deformation mechanism is applied to account for the anatomical changes caused by the spatial normalisation. A set of consistent landmarks spanning over the entire cortical surface were delineated in [210] to compare the performance between landmark-based and automatic surface registration methods. It has been found that the landmark-based method is more reliable to avoid crude registration errors. A boundary-based registration method proposed in [121] extracts tissue boundaries from the static image with high resolution and quality. The lower quality image is aligned to the static image by maximising the intensity gradient across tissue boundaries. It was demonstrated that this method is robust to intensity inhomogeneity and capable of matching partial brain images to whole brain images. A diffeomorphic sulcal-based cortical (DISCO) technique [24] automatically extracts sulcal features from brain images. The diffeomorphic transform is estimated to align the sulcal features across individuals. The authors stated that DISCO can be efficiently combined with an image based deformation to further improve the registration performance. A large deformation diffeomophic registration algorithm was employed in [147] to align brain images by establishing multistructure correspondences through concurrent subcortical and cortical shape matching. This work demonstrated that the incorporation of multiple structure segmentations improves local registration accuracy at cost of the large amount of computational resources. An approach combining volumetric and surface (CVS) registration was proposed in [221]. The registration is performed in two steps: a surface-based registration relying on the extracted geometric features and an optical flow registration to further refine the alignment of noncortical structures. This method establishes the correspondence across the entire brain that accurately aligns both cortical and noncortical regions. It outperforms HAMMER both in registration accuracy and robustness. The integration of sulcal and gyral curves, cortical surfaces, and images is adopted for whole brain diffeomorphic metric mapping in [94]. The diffeomorphic transformation that brings one brain to another is estimated by combining a shape space of intensity images and point set into a unified framework. This method outperforms CVS and HAMMER in terms of cortical and subcortical volume segmentation. Based on HAMMER registration, a symmetric diffeomorphic registration algorithm [319] was developed, which integrates

the hierarchical attribute matching and symmetric diffeomorphic deformation. The registration starts simultaneously from both the moving image and the static image until they meet at the middle point. In contrast to the performance of 14 nonlinear registration algorithms described in [151] on the same datasets, this method consistently generates the best registration results.

5.1.2 Our Work

Demons is an intensity-based, non-rigid registration method proposed by Thirion [281, 282]. It offers a good tradeoff between the registration accuracy and the execution speed. It is popular due to its fast convergence and simple implementation [53]. Most demons methods use the whole image or single scale edges as demon points. Single scale edges are extracted by setting an intensity gradient threshold or applying some edge detection methods [157]. In this work, we present a demons method using salient edges detected in scale-space. 3D scale-space edges are detected from 3D images by extending Lindeberg's scale-space theory [170] into 3D space. These edges are then used as demons to drive non-rigid registration. The impact of edge saliency is investigated by implementing experiments on a 3D MR image database. The registration accuracy is measured using three metrics, demonstrating that salient edges are more important in demons registration than either non-salient edges or randomly sampled points. The performance of the demons method using a certain number of salient edges at three low resolutions is comparable with using all points at four resolutions while the execution time is dramatically reduced.

5.2 Demons registration

The principle of demons originated from Maxwell's thermodynamic theory, which separates a mixture of two types of particles automatically using a semi-permeable membrane. When it is applied to image registration, the contour points in the static image are considered as semi-permeable. And the content of the moving image moves through or away from the demon points on the contour according to the driving force calculated at these points.

Demons based registration is distinct from attractor based registration in several aspects. For attractor based method, the points in the moving image are attracted by the points in the static image in terms of distance or similarity. An efficient algorithm is needed to find the 'closest and most similar' [282] point in the static image. And the forces are calculated for the points in the moving image. For a demons based method,

the moving image is filtered by the demon points in the static image. Whether the moving image moves inside or outside depends on the polarity of the driving force calculated at the demon points.

In this section, we will elaborate the procedures of demons registration including the preprocessing steps, demon points selection, driving force calculation and the registration algorithm implementation.

5.2.1 Preprocessing

Intensity correspondence

Due to the difference of image acquisition process, the intensity distribution of a moving image and a static image is usually not the same. Since our registration method is based on intensity, it is necessary to realise the intensity correspondence before the registration. We use a simple polynomial to estimate the intensity transform

$$I_{new} = a_1 + a_2 I_{orig}, \tag{5.1}$$

where I_{orig} is the original intensity, I_{new} is the new intensity value, and a_1, a_2 are the coefficients. By solving this problem, the cluster centers estimated using an EnFCM method [276] are considered as the representative intensity values for background, CSF, GM and WM for both images. Then the coefficients are determined for these values using a linear least squares method.

Affine registration

Demons registration is based on the concept of optical flow which assumes the intensity at the same location remains constant in the images to be matched. Therefore, a partial overlap between the structures in these images is required before applying demons registration. In this work, an affine registration is used to bring two images into a global correspondence and make the following non-rigid registration more efficient. Each voxel $\mathbf{x} = (x_1, x_2, x_3)$ in the moving image is moved to its corresponding position $\mathbf{y} = (y_1, y_2, y_3)$ in the static image according to:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix}. \tag{5.2}$$

The intensity values of all the voxels contribute to estimate the parameters denoted as $m_{ij}(i = 1,2,3; j = 1,2,3,4)$. These parameters determine the transformation of rotation, scaling, translation and shearing between two images. And the gradient steepest decent algorithm [110] is used for the parameter optimisation.

5.2.2 Demons Selection

To select more efficient demons and investigate how demons selection affects the registration performance, we extended the edge detection method based on scale-space theory [170] into 3D. Both the sharp and diffuse edges are extracted and the edge saliency is calculated.

The scale-space (see [170] for details) representation L of an image f is defined by convolving f by a Gaussian filter g

$$L(.;\, t) = g(.;\, t) * f, \tag{5.3}$$

where $g(.;\, t) = \frac{1}{2\pi t} e^{-(x^2+y^2+z^2)/(2t)}$ and t is the scale parameter. Then the scale-space derivatives are obtained from

$$L_{x^\alpha y^\beta z^\gamma}(.;\, t) = \partial_{x^\alpha y^\beta z^\gamma} L(.;\, t) = g_{x^\alpha y^\beta z^\gamma}(.;\, t) * f, \tag{5.4}$$

where α, β and γ denote the order of differentiation in each dimension.

According to scale-space theory, the edge surface at each scale is defined as

$$L_{vv} = 0; \quad L_{vvv} < 0, \tag{5.5}$$

where v denotes the gradient direction of image L. L_{vv} and L_{vvv} are the second-order and third-order directional derivatives in the v-direction. In 3D, this constraint can be restated as

$$\begin{cases} L_{vv} \propto L_v^2 L_{vv} = L_x^2 L_{xx} + L_y^2 L_{yy} + L_z^2 L_{zz} \\ \qquad + 2L_x L_y L_{xy} + 2L_x L_z L_{xz} + 2L_y L_z L_{yz} = 0, \\ L_{vvv} \propto L_v^3 L_{vvv} = L_x^3 L_{xxx} + L_y^3 L_{yyy} + L_z^3 L_{zzz} \\ \qquad + 3L_x L_y^2 L_{xyy} + 3L_x L_z^2 L_{xzz} + 3L_x^2 L_y L_{xxy} \\ \qquad + 3L_x^2 L_z L_{xxz} + 3L_y^2 L_z L_{yyz} + 3L_y L_z^2 L_{yzz} \\ \qquad + 6L_x L_y L_z L_{xyz} < 0. \end{cases} \tag{5.6}$$

After extracting the edge surfaces in scale-space, the edge strength is measured by

$$\varepsilon_{\lambda-norm}L = L^2_{v,\lambda-norm} = t^{\lambda}(L_x^2 + L_y^2 + L_z^2), \tag{5.7}$$

where λ is a parameter that is stated essential when formulating scale-space edge detection mechanisms and is usually set to $1/2$ [170]. Scale-space edges can be found at the intersection of the edge and $\varepsilon_{\lambda-norm}L$ surfaces. This is formulated as

$$\begin{cases} \partial_t(\varepsilon_{\lambda-norm}L(x,y,z;\ t)) = 0, \\ \partial_{tt}(\varepsilon_{\lambda-norm}L(x,y,z;\ t)) < 0, \\ L_{vv}(x,y,z;\ t) = 0, \\ L_{vvv}(x,y,z;\ t) < 0. \end{cases} \tag{5.8}$$

The scale-space edges are then considered as demon points. And the edge saliency is calculated as

$$G = \sqrt{2\pi t(L_x^2 + L_y^2 + L_z^2)}. \tag{5.9}$$

5.2.3 Driving Force

The theory of optical flow is most commonly used to estimate the demons driving force [281]. It works under the assumption that the moving image m and the static image s are two frames of a motion sequence and the intensity of each voxel under motion remains constant. Therefore, the motion vector $\overrightarrow{f}$ that brings m to s is estimated as

$$\overrightarrow{f} = \frac{(m-s)\overrightarrow{\nabla} s}{|\overrightarrow{\nabla} s|^2 + (m-s)^2}, \tag{5.10}$$

where $\overrightarrow{\nabla} s$ represents the gradient of image s. The vector $\overrightarrow{f}$ could be used as the demons driving force because its behaviour meets the requirement for the demon points. The moving image m is pushed in the direction of $\overrightarrow{\nabla} s$ when $s < m$ and in the direction of $-\overrightarrow{\nabla} s$ when $s > m$, as illustrated in Fig. 5.1.

The demons driving force based on the optical flow theory in the classic demons method is calculated using the gradient information of the static image only. Considering this, Wang et al [304] proposed an improved force by adding an 'active' force derived from the gradient information of the moving image. The improved driving

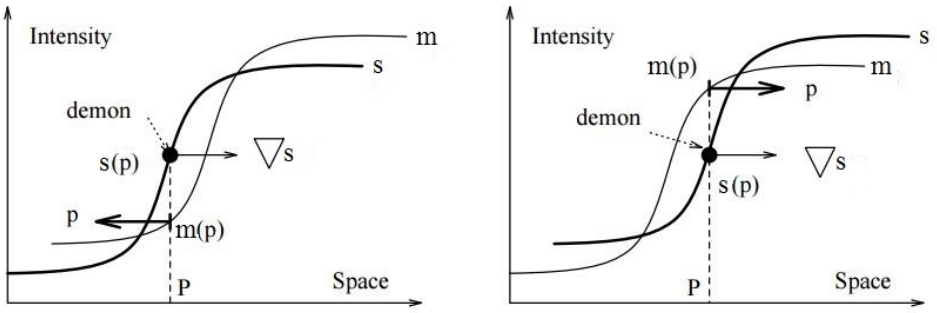


Fig. 5.1 Image m is pushed by the demon point P to different directions according to the intensity values of $m(P)$ and $s(P)$. The original figure is from [281].

force is calculated as

$$\overrightarrow{f} = (m-s)\left(\frac{\overrightarrow{\nabla} s}{|\overrightarrow{\nabla} s|^2 + \omega^2 (m-s)^2} + \frac{\overrightarrow{\nabla} m}{|\overrightarrow{\nabla} m|^2 + \omega^2 (s-m)^2}\right), \tag{5.11}$$

where ω is a normalisation factor proposed by Cashier et al. [52] to adaptively adjust the strength of the force. Then the step size is bounded to $\frac{1}{2\omega}$.

5.2.4 Algorithm Implementation

A multisresolution implementation can speed up the convergence of demons registration and make the algorithm more robust [281]. The images are downsampled into 3 or 4 coarse-resolution images by a factor of 2. The affine transformation is used as the initialisation for the first match between the images at the coarsest resolution. Then the resulting transformation from the demons registration at the coarsest resolution is upsampled to initialise the next match at a finer resolution. Thus the solution is refined from coarse-to-fine resolutions. For the matching at each scale, an iterative scheme is well suited for the demons method. The entire demons registration process is shown in Algorithm 1.

In this algorithm, $\circ$ is the transformation operator; σ is a Gaussian kernel used to regularise the transformation field T. The regularisation is applied to ensure the smoothness of the transformation and propagate the displacements on demon points to the whole image grid. The convergence of the iterative process at each scale is controlled by

$$E = \frac{1}{N} \sum (\|s - m \circ (T + \mathbf{U})\|^2 (1 + \omega^2 \|\mathbf{U}\|^2)), \tag{5.12}$$

Algorithm 1: Demons Algorithm

input : The static image s and the moving image m; The affine transform T_{aff}; The number of resolutions n;
output : The demons transform T;

begin
  T is initialised using T_{aff};
  for $i \leftarrow 1$ **to** n **do**
    T is resized to the current scale i;
    s and m are resized to s_i and m_i;
    $m_i \leftarrow m_i \circ T$;
    Demons selection from s_i;
    repeat
      Computing the update field **U** using the demons driving force defined in Eq. 5.11;
      $T \leftarrow T + \mathbf{U}$;
      T is regularised by convolving with σ;
      $m_i \leftarrow m_i \circ T$;
    until *Convergence*;

where N is the total number of image voxels. The algorithm is assumed to have converged when the difference of E between any pair of the most recent consecutive iterations is under a very small threshold ε.

5.3 Experiments and Results

5.3.1 Database

We used a publicly available database LPBA40 provided by the Laboratory of Neuro Imaging at USC [253] in our experiments. 40 T1-weighted brain MR images in 3D were provided. All the images were skull stripped and 56 structures were manually labelled for each subject. The image size is $181 \times 217 \times 181$.

5.3.2 Evaluation Methods

According to the work in [234], the overlaps of sufficiently local labelled regions of interest, surfaces and lines are more reliable to quantify registration accuracy. In this study, we used three region volume overlap metrics described in Section 3.4.1: the target overlap TO, the mean overlap MO and the false positive rate FPR. These

metrics measure the brain volume overlap by accumulating the overlap between each pair of regions in the moving and static images. TO and MO measure the overlap agreement while FPR measures the overlap error between the moving and static brain images. Paired-sample t-test was used to investigate the impact of parameters on the registration performance of specific methods with the significance level set at 0.05 and the repeated-measures MANOVA was applied to compare the performance of various registration approaches with regard to several metrics.

5.3.3 Experiment Settings

In the experiments, 40 brain volumes were randomly divided into 20 pairs. The moving and static images were also randomly determined in each pair. Each volume was processed only once in an experiment to ensure the independence of 20 sets of registration results. For every pair of images, intensity and global correspondences were completed in preprocessing. Then the demons registration was applied to obtain the non-rigid transform between them. In the evaluation, 56 region masks for both of the moving and static images were generated from their structure labels. Then each of the moving region masks was registered by applying the affine and non-rigid transforms (with linear interpolation). Finally, registration performance was evaluated by comparing the registered and static region masks based on the metrics described above.

5.3.4 Whole Image Grid As Demon Points

In the first phase of the experiments, all points in the images were used as demon points. Demons registration was applied in a multiscale implementation with image resolutions of 0.125, 0.25, 0.5 and 1. We analysed the impact of the normalisation factor ω in the improved driving force defined in Eq. 5.11.

Table 5.1 Impact of ω in the improved driving force on demons registration performance measured by FPR

	0.5	1	2	2.5	3	3.5	4
0.5	$\mathbf{0.2879 \pm 0.0692}$	0.0059*	0.0176*	0.0315*	0.0508	0.0757	0.1091
1	–	$\mathbf{0.2761 \pm 0.0557}$	0.0827	0.1553	0.2657	0.4029	0.5728
2	–	–	$\mathbf{0.2709 \pm 0.0450}$	0.5773	0.9773	0.6347	0.3780
2.5	–	–	–	$\mathbf{0.2703 \pm 0.0407}$	0.2357	0.1065	0.0449*
3	–	–	–	–	$\mathbf{0.2710 \pm 0.0387}$	0.0351*	0.0122*
3.5	–	–	–	–	–	$\mathbf{0.2720 \pm 0.0372}$	0.0046*
4	–	–	–	–	–	–	$\mathbf{0.2732 \pm 0.0359}$

- The numbers in the first column and the first row are the ω values tested for demons using the improve driving force on LPBA40. The values highlighted in bold show the mean FPR values and the standard deviations produced by demons with the corresponding ω. Each of the other numbers in the table shows the p-value tested between two sets of registration results produced with a pair of ω values in the corresponding row and column. The significant differences are indicated by $\star$ at the top right of the p-values.
- The p-values in the table are symmetric, so we only show half of them and the omitted ones are replaced by '–'.

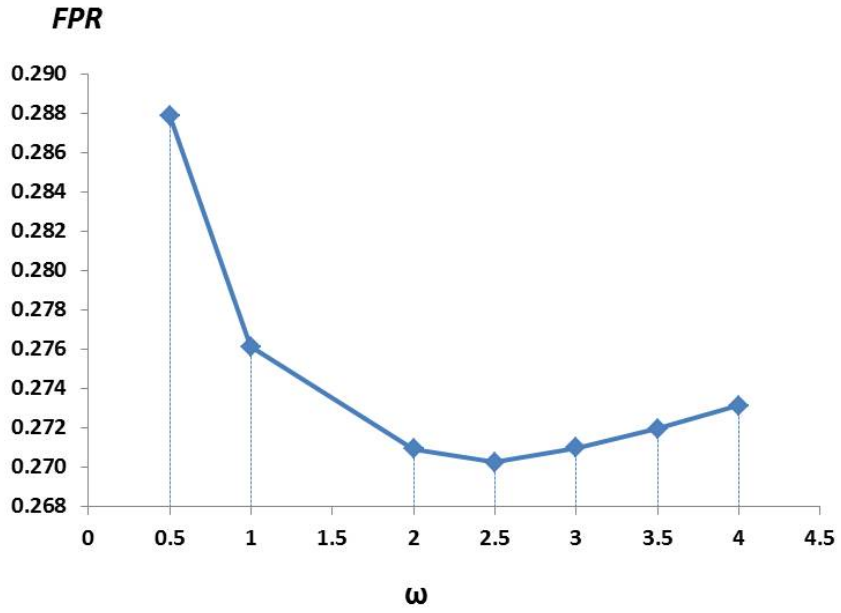


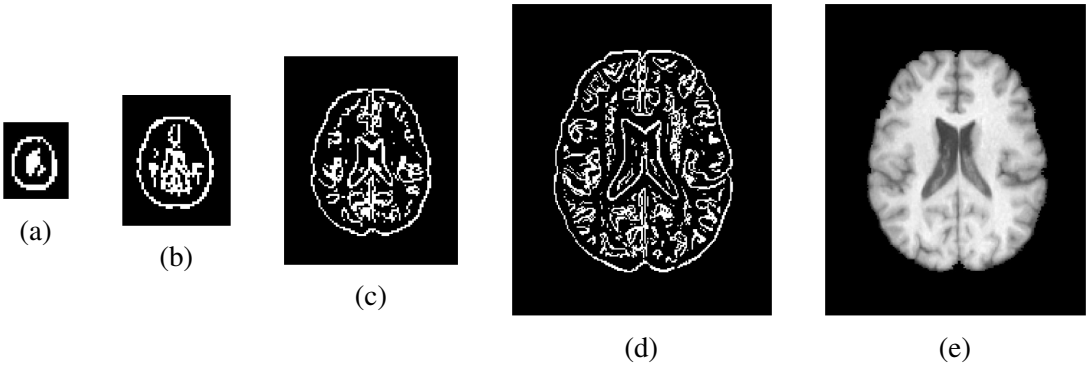
Fig. 5.2 Mean FPR values when varying the parameter ω in the improved driving force used for demons registration.

Impact of ω in the improved driving force

The impact of ω in the improved driving force on demons registration is analysed in Fig. 5.2. A set of ω values were tested and the registration performance was measured by FPR. The paired-sample t-test was adopted to test the difference between any two sets of results generated by demons with a pair of ω values. The FPR value produced with each ω and the p-values calculated from applying the t-test are listed in Table 5.1. ω starts from 0.5 and gradually increases. The performance of demons registration is significantly improved (the FPR value is decreased) when ω increases to 1. Then no significant differences are found when gradually increasing ω until 3. After that point, the registration is significantly deteriorated when further increasing the value of ω. Hence the optimal value of ω is between 1 and 3. No significant difference exists when selecting any value in this range. In the following experiments, ω is set to 1.

5.3.5 Edges As Demon Points

In the second phase of our experiments, instead of using all points in the images, we used all the edges detected by applying the method described in Section 5.2.2. Three scales 1, 2 and 3 were used for the scale parameter t to select the demon points and the registration performance was compared with that of using the whole image as



(a) (b) (c) (d) (e)

Fig. 5.3 Images of all the edges detected at different resolutions: resolution 0.125 (a); resolution 0.25 (b); resolution 0.5 (c) and resolution 1 (d). The original image is shown in (e).

demons. Fig. 5.3 illustrates the images of all the edges detected at four resolutions. Table 5.2 lists the mean brain volume overlaps with the standard deviations calculated before registration and after applying affine registration, the demons method using all points and the demons method using edges only. Repeated-measures MANOVA was used to test the differences of brain volume overlaps among four different states. The multivariate outcome tells significant differences exist among brain volume overlaps in four different states. The univariate outcomes demonstrate that significant differences lie among the four states in terms of all three metrics. Then the ***post hoc*** pairwise comparisons are conducted between brain volume overlaps in any two different states measured by TO, MP and FPR, respectively. It is observed that the affine registration significantly improves the brain volume overlaps compared with those before registration in terms of all three metrics; the demons method using either all points or the edges only further significantly improves the brain volume overlaps compared with those after affine registration with respect to TO; the demons method using the edges only performs comparatively with that using all points in terms of all three metrics.

5.3.6 Salient Edges As Demon Points

Considering different saliencies of edges, we used a certain number of edges with high saliency (see Fig. 5.4) instead of using all of them for demons registration. The performance of salient edge guided demons registration applied at four and three resolutions is analysed.

Table 5.2 Brain volume overlaps before and after registration

	TO	MO	FPR
1	0.5981 ± 0.0675	0.5990 ± 0.0558	0.3933 ± 0.0691
$2^{\star}$	$0.6828 \pm 0.0198^{\star}$	$0.6876 \pm 0.0133^{\star}$	$0.3072 \pm 0.0162^{\star}$
$3^{\star\star}$	$0.7009 \pm 0.0824^{\star\star}$	$0.7111 \pm 0.0670^{\star}$	$0.2761 \pm 0.0557^{\star}$
$4^{\star\star}$	$0.7173 \pm 0.0463^{\star\star}$	$0.7128 \pm 0.0435^{\star}$	$0.2914 \pm 0.0435^{\star}$

- Each value represents the mean brain volume overlap and the standard deviation before or after applying registration on LPBA40. The numbers in the first column represent different states of brain volumes (1: before registration; 2: after affine registration; 3: after demons registration using all points; 4: after demons registration using the edges only).
- The $\star$s at the top right of each value indicate the significant differences of the corresponding brain volume overlap compared with all the others in other states measured by the same metric. The $\star$s at the top right of each state number indicate the significant differences of the overall brain volume overlap in the corresponding state compared with all the other states. The brain volume overlap or the state labelled with more $\star$s is significantly better than those with fewer or no $\star$s and the brain volume overlaps or the states labelled with equal number of $\star$s have no significant difference between each other.

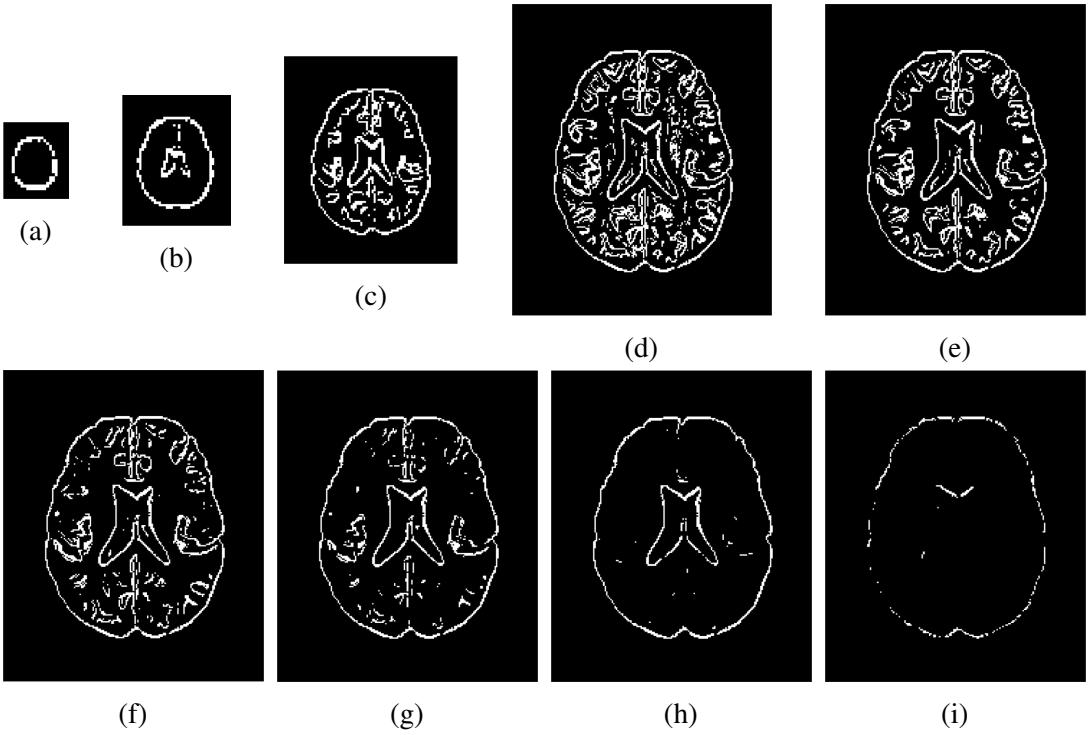


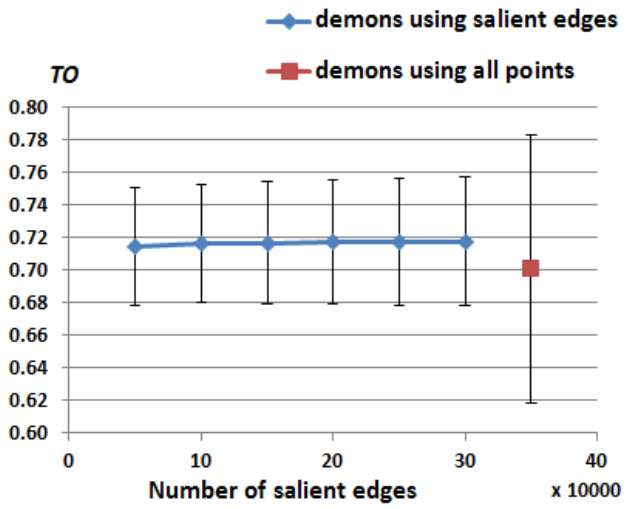
Fig. 5.4 Images of salient edges detected at each resolution: resolution 0.125 (a); resolution 0.25 (b); resolution 0.5 (c); resolution 1 with 300,000 edges (d), 250,000 edges (e), 200,000 edges (f), 150,000 edges (g); 100,000 edges (h) and 50,000 edges (i).

Demons registration at four resolutions

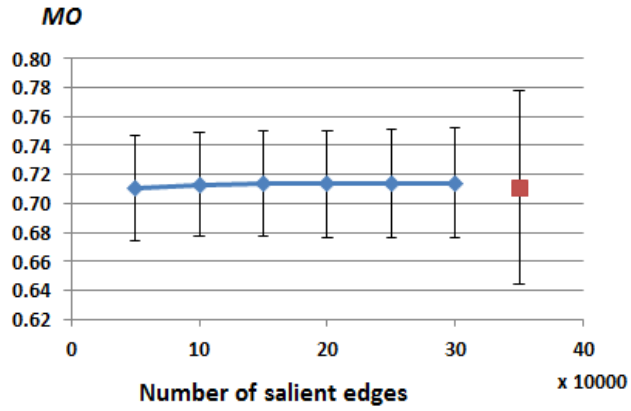
Most of the edges detected at low resolutions were used in this experiment (900, 5000 and 40000 edges at resolutions 0.125, 0.25 and 0.5). At the original resolution 1, the number of edges was set to 300000 first and then decreased by 50000 until 50000. By applying the paired-sample t-test, we found that no significant differences exist when decreasing the number of salient edges compared with that of using all points as demon points as shown in Table 5.3. Fig. 5.5 illustrates the mean accuracies and the standard deviations generated by demons using various number of edges as demon points, measured by TO, MO and FPR, respectively.

Demons registration at three resolutions

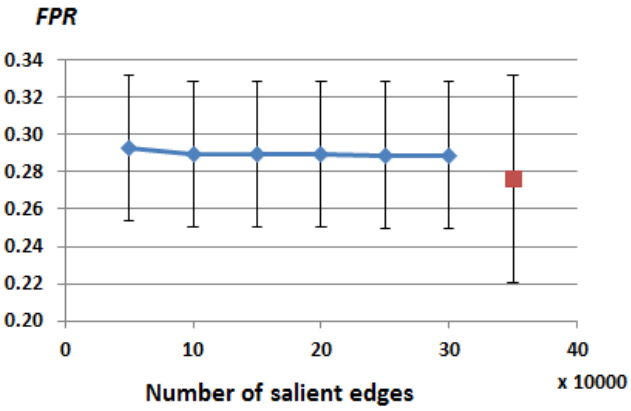
Then we considered using only three coarse resolutions 0.125, 0.25 and 0.5 for demons registration. The numbers of salient edges at resolutions 0.125 and 0.25 were kept the same. We reduced the number of edges at resolution 0.5 from 40000 to 10000 at a step size of 10000. Similarly, the paired-sample t-test was used to test the difference of



(a)



(b)



(c)

Fig. 5.5 Comparisons of demons registration performance between using various number of salient edges at resolution 1 and the whole grid of image as demon points. The registration performance is measured by TO (a), MO (b) and FPR (c), respectively, with the corresponding standard deviation shown for each result.

Table 5.3 Comparisons of the performance of demons registration between using salient edges and all points

	300,000	250,000	200,000	150,000	100,000	50,000
TO	0.1352	0.1463	0.1614	0.1745	0.1917	0.2612
MO	0.7109	0.7338	0.7642	0.7814	0.8189	0.9601
FPR	0.1564	0.1555	0.1562	0.1698	0.1707	0.1040

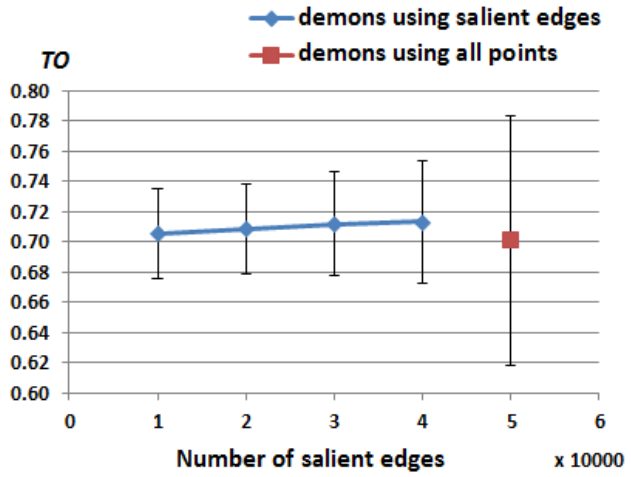
- The numbers in the top row represent the numbers of salient edges used as demon points at resolution 1.
- Each value represents the p-value between the registration accuracies generated by demons using the corresponding number of salient edges and the whole grid of image as demon points with regard to each metric.

demons registration performance between using all points at 4 resolutions and a certain number of salient edges at resolution 0.5 as demon points. The p-values are listed in Table 5.4. Fig. 5.6 illustrates the demons registration performance using various number of salient edges as demon points at resolution 0.5 with respect to each metric. It is shown that in terms of TO and MO, no significant differences exist between registration accuracies generated by demons using all points at four resolutions and a certain number of salient edges at three resolutions; in terms of FPR, no significant difference is found until the number of salient edges at resolution 0.5 is reduced to 10,000.

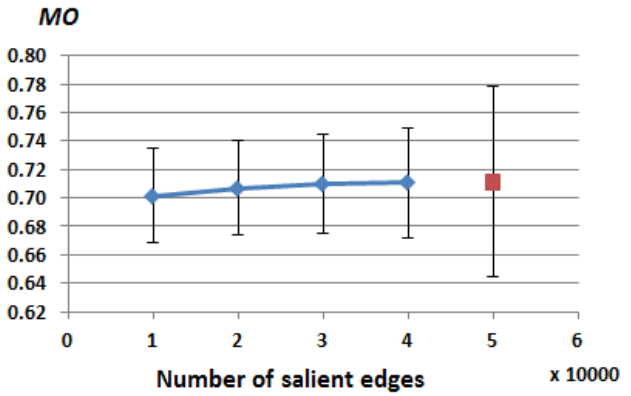
Table 5.4 Comparisons of the performance of demons registration between using salient edges and all points

	40,000	30,000	20,000	10,000
TO	0.2594	0.3748	0.5754	0.7382
MO	0.9486	0.9228	0.7028	0.3706
FPR	0.0808	0.1460	0.1192	$0.0349^{\star}$

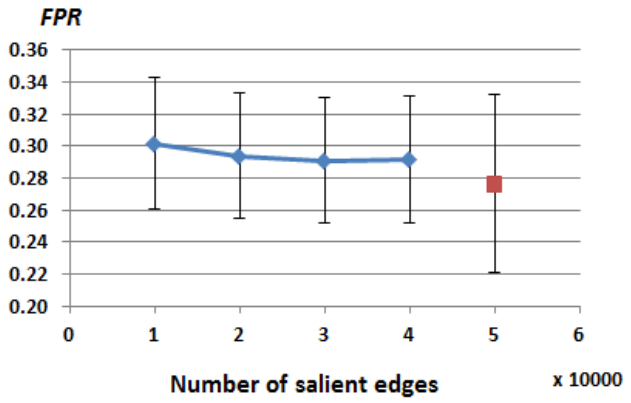
- The numbers in the top row represent the numbers of salient edges used as demon points at resolution 0.5.
- Each value represents the p-value between the registration accuracies generated by demons using the corresponding number of salient edges and the whole grid of image at 4 resolutions as demon points with regard to each metric. The significant difference is indicated by $\star$.



(a)



(b)



(c)

Fig. 5.6 Comparisons of demons registration performance between using various number of salient edges at resolutions 0.5 and the whole grid of image as demon points. The registration performance is measured by TO (a), MO (b) and FPR (c), respectively, with the corresponding standard deviation shown for each result.

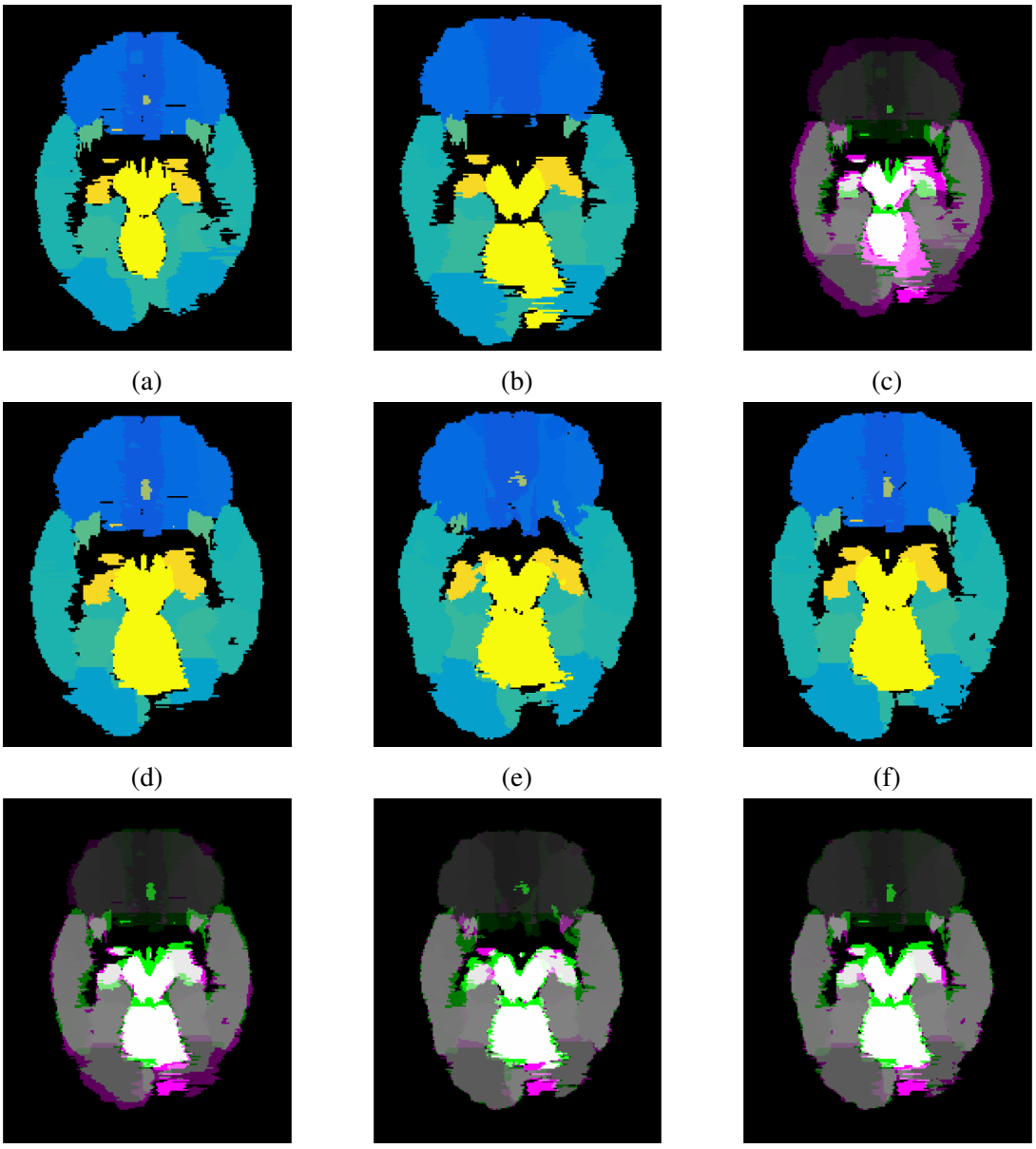


Fig. 5.7 Qualitative comparisons of three different registration approaches. (a) and (b) are label maps of axial slices from the moving image and the static image; (d-f) are the corresponding label maps of (a) by applying affine registration, demons registration using all points and demons registration using salient edges between the moving image and the static image; (c) and (g-i) show the overlaps between the label maps of the moving image before and after registration shown in (a, d-f) and the label map of the static image (b). For the label maps (a, b, d-f), different colours represent different brain regions; for the images representing the overlaps (c, g-i), green and magenta regions show the differences between two label maps. The brain overlaps between the moving image and the static image before and after various registration approaches are 0.4332, 0.6759, 0.7153 and 0.7123 measured by TO; 0.4995, 0.6893, 0.6982 and 0.7129 measured by MO; 0.4101, 0.2968, 0.3182 and 0.2866 measured by FPR.

Fig. 5.7 gives qualitative comparisons of the registration performance of applying affine registration (Fig. 5.7(d, g)), demons method using all points (Fig. 5.7(e, h)) and demons method using salient edges (Fig. 5.7(f, i)). The demons method using all points is applied at four resolutions and the demons method using salient edges is applied at three resolutions with 20000 edges used at resolution 0.5. From Fig. 5.7(c), we can see that the moving image has a poor correspondence with the static image before registration; it is generally aligned with the static image in size and shape by applying the affine registration (Fig. 5.7(g)); then the demons registration using all points at 4 resolutions further refines the registered image most obviously at the boundaries (Fig. 5.7(h)); the demons method using salient edges only at 3 resolutions (Fig. 5.7(i)) also corrects the deformed image generated by the affine registration at the boundaries and it performs comparatively with that using all points at 4 resolutions.

This experiment shows that by contrast to using the whole grid of image, using salient edges can reduce the number of demon points and resolution levels without deteriorating the registration performance significantly. Apart from the preprocessing steps, the running time of demons registration process was substantially reduced to $78.56 \pm 1.74s$ (when using 3 resolutions and the number of edges at resolution 0.5 was 20000) from $246.94 \pm 15.01s$ (when using all points at 4 resolutions).

5.3.7 Impact of Demons Selection

To demonstrate the importance of salient edges over other points in demons registration, we compared the performance of using salient edges, non-salient edges, randomly sampled edges and randomly sampled points. Fig. 5.8 gives an example of these demons selection mechanisms and their corresponding registration accuracies measured by each metric are illustrated in Fig. 5.9. For demons methods not using all points (method sedg3, nsedg3, redg3 and rpt3 in Fig. 5.9), three resolutions (0.125, 0.25 and 0.5) were used and the numbers of demon points were set to 900, 5000 and 20000, respectively. Paired-sample t-test was applied to test the difference between the registration performance produced by demons applied at four (method all4) and three resolutions (method sedg3, nsedg3, redg3, rpt3 and all3) with respect to each metric. The comparisons listed in Table 5.5 indicate that by contrast to the method using all points at 4 resolutions (method all4), all the other methods using 3 resolutions (method nsedg3, redg3, rpt3 and all3) perform significantly worse except the method using salient edges (method sedg3). Further more, from the results listed in Table 5.6, it shows that registration using salient edges (method sedg3) obtains significantly

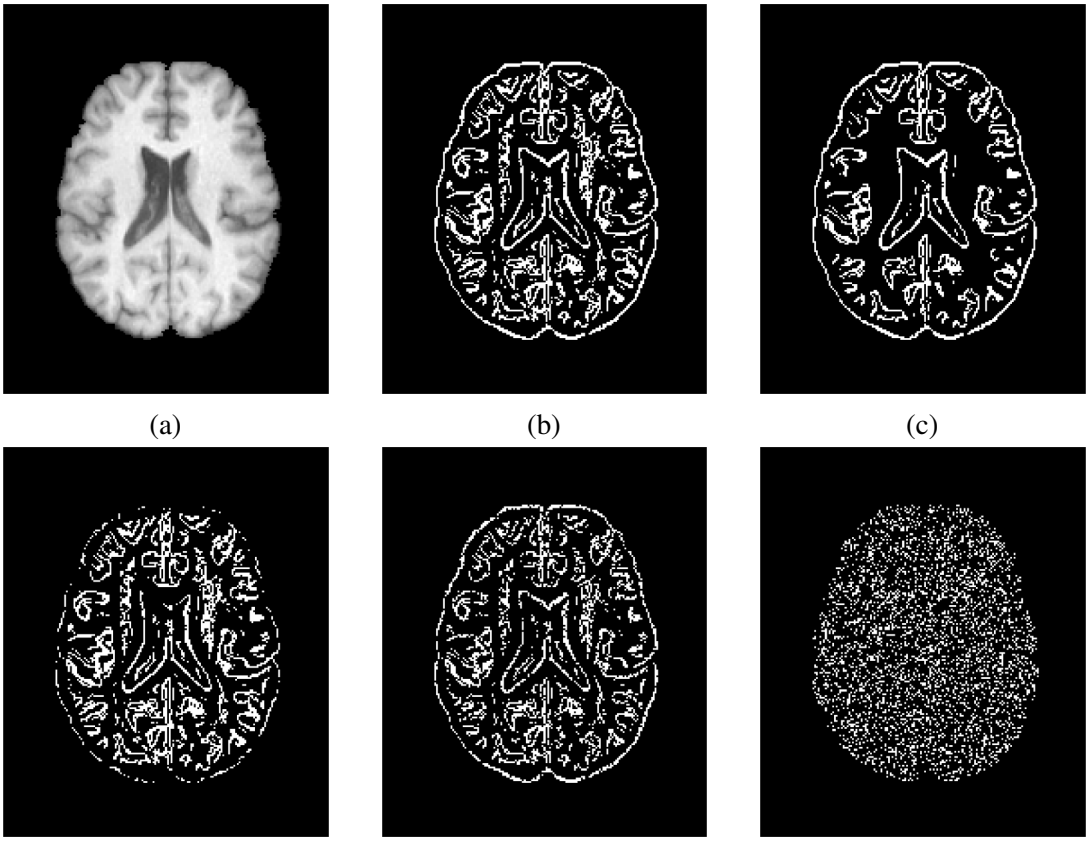
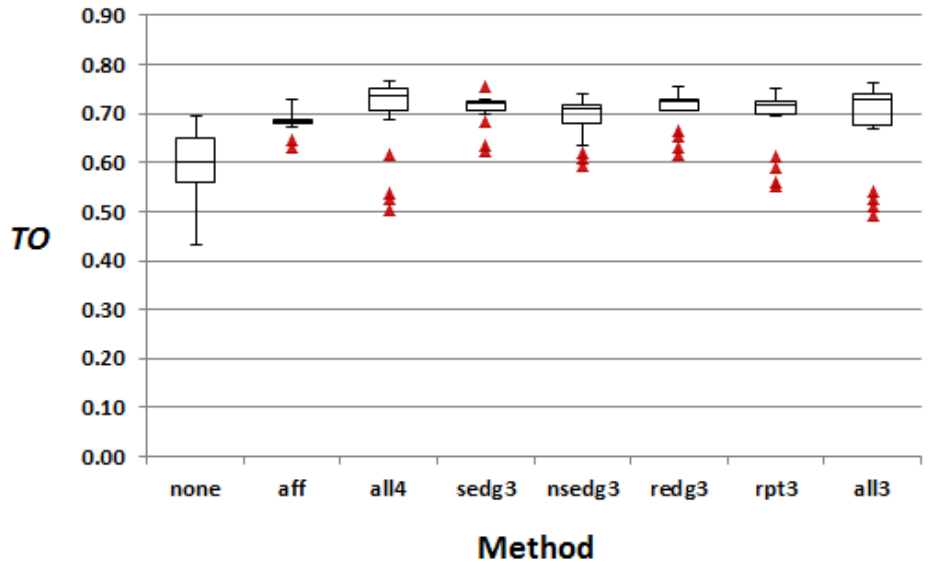


Fig. 5.8 Various ways of demon points selection in an axial brain slice: the original image (a); all the detected edges (b); and the same number of salient edges (c), non-salient edges (d), randomly sampled edges (e) and randomly sampled points (f).

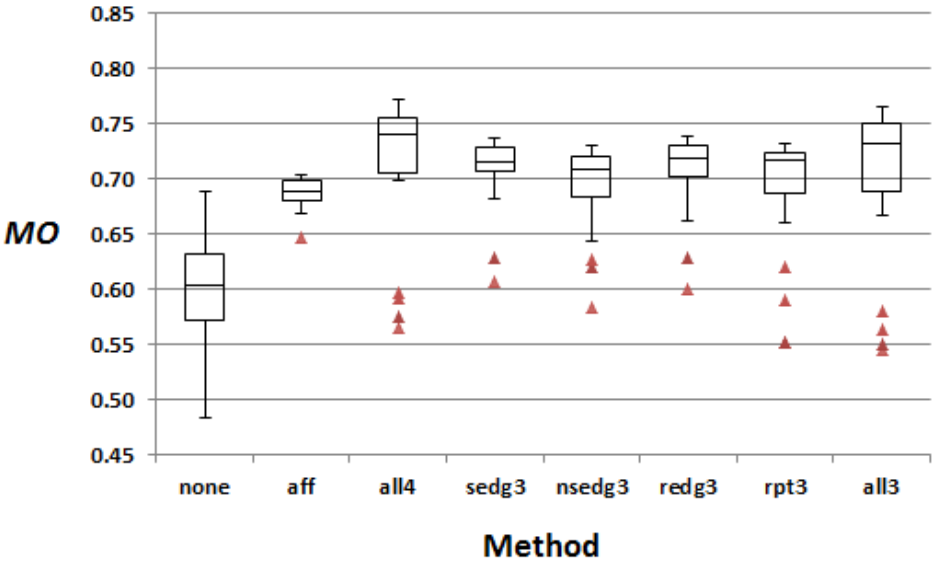
better accuracy than that using either non-salient edges (method nsedg3) or randomly sampled points (method rpt3).

5.4 Conclusions

This chapter presented a demons method using scale-space edges as demon points. We have investigated the impact of edge saliency in demons registration performance. A series of experiments have been conducted on a publicly available database and the registration accuracies have been compared qualitatively and quantitatively using three region based metrics. The results show that salient edges contribute more than either non-salient edges or randomly sampled points in demons registration. Using three low resolutions with a certain number of salient edges as demons at each resolution, statistically equal performance is obtained compared with using all points at 4 resolutions. And the execution time is dramatically reduced by nearly $2/3$.

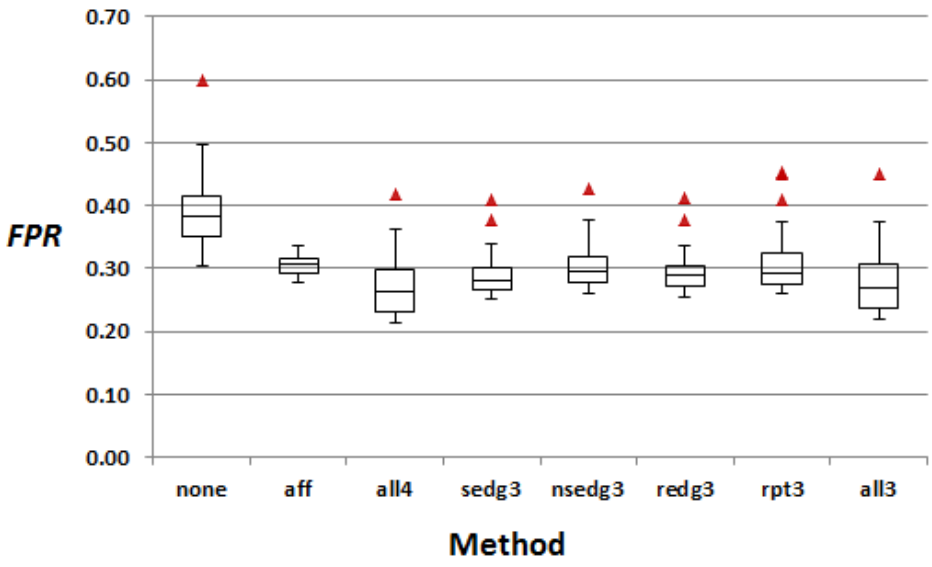


(a)



(b)

Fig. 5.9 Comparisons of demons registration performance based on various demons selection mechanisms.



(c)

Fig. 5.9 Comparisons of demons registration performance based on various demons selection mechanisms continued. The registration performance is measured by TO (a), MO (b) and FPR (c), respectively. The numbers on the horizontal axes represent different status under which the brain overlaps are measured: before registration (none); after affine registration (aff); after demons registration using all points at 4 resolutions (all4); and after demons registration using salient edges (sedg3), non-salient edges (nsedg3), randomly sampled edges (redg3), randomly sampled points (rpt3) and all points (all3) at 3 resolutions. Each box represents the registration result produced by a method and has lines at the lower quartile, median and upper quartile values; whiskers extend from the ends of the box to the most extreme values within the inner fences (1.5 times the interquartile range from the ends of the box). Outliers with values beyond the fences are marked as '▲'.

Table 5.5 Comparisons of the performance of demons registration applied at four and three resolutions

	sedg3	nsedg3	redg3	rpt3	all3
TO	0.5754	0.4040	0.4570	0.1682	$0.0010^{\star}$
MO	0.7028	$0.0146^{\star}$	0.5512	$5.32e-06^{\star}$	$2.29e-04^{\star}$
FPR	0.1192	$5.42e-04^{\star}$	$0.0291^{\star}$	$1.19e-04^{\star}$	$1.49e-05^{\star}$

- The top row represent different demons registration methods, which are consistent with those in Fig. 5.9.
- Each value represents the p-value between the registration accuracies generated by demons applied at four resolutions (method all4) and the corresponding method with regard to each metric. The significant difference is indicated by $\star$.

Table 5.6 Comparisons of the performance of demons registration between using salient edges and other points

	nsedg3	rpt3
TO	$0.0023^{\star}$	0.0560
MO	$0.0039^{\star}$	$0.0192^{\star}$
FPR	$0.0097^{\star}$	$0.0063^{\star}$

- The top row represents different demons registration methods, which are consistent with those in Fig. 5.9.
- Each value represents the p-value between the registration accuracies generated by demons using salient edges (sedg3) and that using non-salient edges (nsedg3) or randomly sampled points (rpt3) at three resolutions with regard to each metric. The significant difference is indicated by $\star$.

Chapter 6

Brain Tissue Classification

6.1 Introduction

Brain tissue classification aims to segment brain tissues, in our case the three primary tissue types: white matter (WM), grey matter (GM) and cerebral spinal fluid (CSF). Such segmentation is essential for diagnosing some brain disorders such as Alzheimer or schizophrenia by quantitative analysis [298, 264]. It can also assist with some applications in medical image analysis like image registration [128], lesion segmentation [143] and cortical surface extraction [149]. T1-weighted magnetic resonance (MR) imaging has been widely used in this area owing to its excellent soft tissue contrast in images [80, 286]. Manual delineation of brain tissue is time-consuming due to the large volume of data. In addition, manual delineation can lead to intra- and inter-expert variability [302]. Numerous supervised and unsupervised algorithms have been developed in the last decade [290, 327, 168]. However, accurate and robust tissue classification remains a challenging task due to noise, intensity inhomogeneity and partial volume effects existing in brain MR images [80, 226, 238].

6.1.1 Previous Work

Different image information priors can be adopted to drive the tissue classification process. Using the features at a voxel level is the most intuitive approach. For example, the intensity information is widely used in clustering methods such as k-Means (KM) [2], Fuzzy c-means (FCM) [329] and Gaussian mixture modelling (GMM) [240, 120]. GMM is more often used by assuming that the intensity of each tissue class follows a Gaussian distribution. The parameters of GMM can be estimated by applying expectation-maximisation (EM) approaches [155] or genetic algorithm [285]. More

advanced voxel-wise features, such as Gaussian scale-space features and Gaussian derivatives [295], and spatial features, such as voxel coordinates [9, 10], can also be extracted from the brain MR images and fed into classifiers to perform tissue classification [301, 294, 307, 199].

Considering the continuity of each tissue type in brain images, the central voxel and its neighbours in a local neighbourhood tend to belong to the same tissue class. This spatial information can be modelled by a Markov random field model (MRF) [80, 326, 286, 109], a hidden Markov random field model (HMRF) [332] or a conditional random field model (CRF) [138] and incorporated into the classification framework as a constraint to improve the accuracy especially in cases where severe image noise and intensity inhomogeneity exist. Besides, features like intensity and location differences with the neighbouring voxels [256] and the normalised means and variance of the neighbouring voxels [283] can be used to benefit the tissue classification of the central voxel.

In addition to the features obtained from a single target image, population-specific atlases can also be introduced as a prior to benefit the classification. This process requires the registration between the atlases and the target image. In the simplest case, the classification is achieved by propagating the label map of a topological atlas to the target image [51]. A topological atlas can also be used to preserve the topology of tissues during classification [33]. To account for the inter-subject variability, a probabilistic atlas can be generated by averaging a series of label maps of other images. These probabilities can be directly utilised as features in the classification framework [286, 34, 323, 222]; or the samples at locations with high probabilities are taken from the registered training images to train a classifier to perform the tissue classification of the target image [83, 300, 314, 66]. In a few cases, features extracted from the probability map, such as Gaussian derivatives [199] and 3D Haar-like features [307], are used for the tissue classification. Both topological and probabilistic atlases are adopted in [34] to segment brain tissues. The topological atlas maintains the topology of brain tissues and the probabilistic atlas initialises the tissue probabilities in the target image. Recently, multi-atlas segmentation (MAS) has drawn great attention because of its superior performance [51, 136, 173, 248]. Instead of using a probabilistic atlas constructed from all the available label maps, a selection process can be applied to select the more relevant atlases, hence not biasing the classification by the less relevant atlases. Atlas selection is carried out based on image similarity or meta-information [6]. After that, the selected atlases are combined using global or local label fusion methods for the final tissue classification [136, 21, 305, 243, 320, 25]. Information from multiple training images or atlases can also be used in patch-based

methods for brain tissue classification. In the work presented in [73], for a target subject, patches are extracted from multi-modality brain MR images. Similar patches are retrieved from the training images and their label maps are combined to realise the tissue classification of the target image. The method proposed in [139] uses the patches extracted from registered multi-modality images to train a random forest. The resulting learned patch ensemble of decision trees are applied to segment brain tissues in the target images. The sparse patch representation is adopted in [306] to fuse multi-modality information and achieve the initial tissue segmentation. The segmentation is iteratively refined by incorporating the anatomical constraint, which further considers the similarity between the segmentations of the training and target images.

Voxel-wise features, neighbouring information and atlas prior information can be combined into an entire framework for brain tissue classification. An approach was proposed in [216], which extracts image intensity features from the central voxel and its neighbours, the posterior probabilities resulting from applying GMM and MRF models and the atlas probabilities produced by registering training images. These image information priors are used to train random decision forests for brain tissue classification. A sub-volume probabilistic atlas is used in [286] to divide the whole image domain into sub-regions with different intensity statistics. Then for each region, the image intensity is modelled by GMM and combined with MRF based regularisation. Local models of all sub-regions are combined into a global probabilistic model to perform the tissue classification. The parameters involved in the algorithm are estimated by applying a genetic algorithm. Image intensity modelled by GMM, MRF based regularisation and atlas priors are combined into an EM approach in [181]. In an approach proposed in [243], the visual appearance descriptor, which models the image intensity as linear combination of discrete Gaussians, and the probabilistic shape prior generated from a subset of co-registered training images are used to achieve the initial tissue segmentation. Subsequently, the segmentation is iteratively refined using a 3D Markov-Gibbs Random Field (MGRF) model, which is described as the spatially invariant second order homogeneity descriptor. A method presented in [248] combines two segmentations based on image intensity and atlas priors using the nearest neighbour transform.

6.1.2 Our Work

The purpose of this chapter is to validate and compare the effects of the three image information priors, including image intensity, neighbouring information and atlas prior, described above on the tissue classification in normal brains. To achieve this, the

performance of four methods modelling them is compared: the image intensity is utilised by KM to obtain a preliminary classification; the neighbouring information is modelled by a MRF and combined with KM by modelling each resulted tissue cluster with a Gaussian distribution; the multi-atlas prior is used in MAS by applying multi-atlas registration and global/local label fusion techniques; finally, we combined the multi-atlas prior with the image intensity and local prior into an overall KM-MRF-MAS framework. In MAS, the importance of the atlas selection is validated and the classification accuracies of applying various label fusion schemes are compared. Finally, we combine the multi-atlas prior with the image intensity and local prior into an overall framework. The effects of a range of parameters in each model are analysed in detail and the performance of all these approaches are compared quantitatively and qualitatively on both simulated and real databases.

6.2 Methods

In this section, we will formulate the modelling of local [80] and multi-atlas priors [25] and the optimisation of tissue classification models constructed using the three image information priors [173, 292].

6.2.1 Problem Formulation

The voxels of a 3D brain MR image are indexed with $i \in \mathscr{S} = \{1,2,...,N\}$ where N is the number of voxels. Each voxel in $\mathscr{S}$ is associated with $y_i \in \mathbb{R}$, the intensity value of the i^{th} voxel. The set of y_i is the observed image denoted by $y = \{y_1, y_2, ..., y_N\} \in \mathbb{R}^N$. The aim of our tissue classification in normal brains is to classify each voxel in $\mathscr{S}$ into one of the classes labelled by $\mathscr{L}$ = {CSF, GM, WM}. The tissue class of the i^{th} voxel is denoted by $x_i \in \mathscr{L}$ and $x = \{x_1, x_2, ..., x_N\} \in \mathscr{L}^N$ is a classification of the image. Our task is to find the best classification $x^* = \{x_1^*, x_2^*, ..., x_N^*\} \in \mathscr{L}^N$ given the image intensity y which can be formulated as a maximum a posterior (MAP) optimisation problem:

$$\begin{aligned} x^* &= \arg\max_{x \in \mathscr{L}^N} P(x|y) = \frac{P(y|x)P(x)}{P(y)} \\ &= \arg\max_{x \in \mathscr{L}^N} P(y|x)P(x) \end{aligned} \tag{6.1}$$

where $P(x|y)$ is the probability of the classification x given the image intensity y. $P(y)$ is independent from the classification x and hence ignored in the optimisation. By taking the negative logarithm of Eq. 6.1, the probability maximisation problem is

converted to finding the minimum of an energy functional:

$$\begin{aligned} x^* &= \arg\min_{x\in\mathscr{L}^N} (-\ln P(y|x) - \ln P(x)) \\ &= \arg\min_{x\in\mathscr{L}^N} (E_{in}(x) + E_{pr}(x)). \end{aligned} \tag{6.2}$$

E_{in} represents the intensity energy which models the intensity distributions of three tissue classes. It measures how well the current classification x explains the image y. E_{pr} represents the prior knowledge of the classification. In Section 6.3, E_{pr} will model the local prior while in Section 6.5, it will model the information from both the local neighbourhood and the multi-atlas.

6.2.2 Intensity and Neighbouring Information Model: KM-MRF

Intensity Prior Energy

The intensity energy E_{in} is calculated by assuming the intensities of all the voxels in images are independently and identically distributed. The likelihood $P(y|x)$ can be expressed as

$$P(y|x) = \prod_{i\in\mathscr{S}} P(y_i|x_i). \tag{6.3}$$

$P(y_i|x_i)$ is the probability density function of y_i given the tissue class x_i. This probability is usually modelled by a Gaussian function for each tissue class [80]:

$$P(y_i|x_i) = \frac{1}{\sigma_{x_i}\sqrt{2\pi}} \exp\left[\frac{-(y_i-\mu_{x_i})^2}{2\sigma_{x_i}^2}\right], x_i \in \mathscr{L}, \tag{6.4}$$

where the model parameters $\theta_{x_i} = \{\mu_{x_i}, \sigma_{x_i}\}$ are the mean and standard deviation of the Gaussian distribution for the tissue class x_i.

E_{in} can be formulated as

$$\begin{aligned} E_{in}(x) &= -\sum_{i\in\mathscr{S}} \ln P(y_i|x_i) \\ &= \sum_{i\in\mathscr{S}} \left(\frac{(y_i-\mu_{x_i})^2}{2\sigma_{x_i}^2} + \ln\sqrt{2\pi} + \ln\sigma_{x_i} \right). \end{aligned} \tag{6.5}$$

Local Prior Energy

The local prior energy E_{pr} can be modelled by a MRF which assumes that the tissue class of the voxel at $\mathscr{S}_i$ depends only on its local neighbourhood $\mathscr{N}_i$ with $i \notin \mathscr{N}_i$ and

$i \in \mathscr{N}_j \leftrightarrow j \in \mathscr{N}_i$ [80]. A Gibbs distribution is often used to characterize a MRF as

$$P(x) = Z^{-1} \exp[-U(x,\beta)], \tag{6.6}$$

where Z is the normalization factor; β is the spatial parameter and $U(x,\beta)$ represents the energy function. A simplified Potts model is generally used to express the energy function $U(x,\beta)$:

$$U(x|\beta) = \sum_{i \in \mathscr{S}} \left(\frac{\beta}{2} \sum_{j \in \mathscr{N}_i} V_{ij}(x_i, x_j) \right). \tag{6.7}$$

V_{ij} models the local spatial transitions between the voxel at $\mathscr{S}_i$ and its neighbour at $\mathscr{S}_j$ and can be calculated as

$$V_{ij}(x_i, x_j) = \frac{\delta(x_i, x_j)}{d(i,j)}, \tag{6.8}$$

where $d(i,j)$ measures the distance between the two voxels. The weighting function $\delta(x_i, x_j)$ is defined as

$$\delta(x_i, x_j) = \begin{cases} -1 & x_i = x_j \\ +1 & x_i \neq x_j. \end{cases} \tag{6.9}$$

This weighting function constrains the classification process with a reward and penalty mechanism. Then E_{pr} can be formulated as

$$E_{pr} = U(x|\beta) + \ln Z. \tag{6.10}$$

Eq. 6.2 can be rewritten as

$$x^* = \arg\min_{x \in \mathscr{L}^N} \sum_{i \in \mathscr{S}} \left(\frac{(y_i - \mu_{x_i})^2}{2\sigma_{x_i}^2} + \ln \sigma_{x_i} + \frac{\beta}{2} \sum_{j \in \mathscr{N}_i} V_{ij}(x_i, x_j) \right) \tag{6.11}$$

with the constants removed.

Classification Approach

The tissue classification is iterated with parameters estimation in an Expectation Maximization (EM) framework which is generalised in Algorithm 2.

Algorithm 2: Tissue Classification Using an EM Approach

1: Initialisation: For each tissue class $l \in \mathscr{L}$, the parameters $\hat{\theta}_l = \{\hat{\mu}_l, \hat{\sigma}_l\}$ are estimated from the KM classification;
2: Classification: Given the current parameter $\hat{\theta}$, the tissues are classified by minimising the energy functional stated in Eq. 6.11 with the Iterated Conditional Modes (ICM) algorithm [35];
3: Maximization: The parameter $\hat{\theta}$ is updated according to the new classification;
4: Iterate steps 2 and 3 until $\hat{\theta}$ is stable.

Given a classification, the parameter estimations for the tissue class l in steps 1 and 3 in Algorithm 2 are computed as

$$\begin{aligned} \hat{\mu}_l &= \frac{\sum_{i \in \mathscr{S}} y_i \xi_{x_i,l}}{\sum_{i \in \mathscr{S}} \xi_{x_i,l}}; \\ \hat{\sigma}_l{}^2 &= \frac{\sum_{i \in \mathscr{S}} (y_i - \hat{\mu}_l)^2 \xi_{x_i,l}}{\sum_{i \in \mathscr{S}} \xi_{x_i,l}}, \end{aligned} \tag{6.12}$$

where $\xi_{x_i,l}$ is defined as

$$\xi_{x_i,l} = \begin{cases} 1 & x_i = l \\ 0 & x_i \neq l. \end{cases} \tag{6.13}$$

6.2.3 Multi-atlas Segmentation: MAS

In MAS, only the atlas priors (i.e. the segmentations of other images) are used to drive the tissue classification. Firstly, multiple atlas images are nonrigidly registered to the target image using demons registration [281, 304]. Then the similarities between the registered atlas images and the target are measured and only the atlases with high similarities are selected to contribute to the classification. After that, the label maps of the selected atlases are propagated to the target image. Finally, global or local label fusion methods are applied to the transformed label maps to infer the classification of the target image. A variety of label fusion methods have been proposed [136]. We use representative approaches to compare the performance of global and local label fusion techniques in MAS. Other methods such as STAPLE [313] or SIMPLE [162] could be used as alternatives.

Atlas Selection

Mutual information and normalised cross correlation are commonly used for atlas selection [6, 136], and we found both of them applicable for atlas selection. We used mutual information [250, 74] to calculate the similarity between the registered atlas a and the target image y:

$$I(a;y) = H(a) + H(y) - H(a,y). \tag{6.14}$$

$H(a)$ and $H(y)$ denote the entropy of a and y, respectively; $H(a,y)$ denotes the joint entropy of a and y. The atlases with high mutual information values are selected for label fusion.

In the following, the set of the selected atlases are denoted by $\mathscr{A}$ with the corresponding label maps $\{x^a | a \in \mathscr{A}\}$. The atlases and the label maps have already been transformed to the target coordinate system. The similarities between the atlases $\mathscr{A}$ and the target image y are denoted by $\{I(a;y) | a \in \mathscr{A}\}$.

Global Label Fusion

Global label fusion methods estimate the classification accuracy of each atlas globally and assign a single weight to each of the selected atlases.

Majority Voting (MV) is the most commonly used label fusion strategy [136, 16]. It treats every atlas equally and assigns to each voxel the class label that most atlases agree on. This process can be formulated as

$$x_i^* = \arg\max_{l \in \mathscr{L}} \sum_{a \in \mathscr{A}} \xi_{x_i^a, l}, \tag{6.15}$$

where x_i^a is the label at the location i in the label map x^a. The function ξ is defined in Eq. 6.13.

Mutual Information (MI) takes image intensities into account and assigns to each atlas a a global weight which is proportional to $I(a;y)$ and normalised by $I_{max} = \max\{I(a;y) | a \in \mathscr{A}\}$:

$$x_i^* = \arg\max_{l \in \mathscr{L}} \sum_{a \in \mathscr{A}} \frac{I(a;y)}{I_{max}} \xi_{x_i^a, l}. \tag{6.16}$$

Local Label Fusion

Local label fusion methods estimate the classification accuracy at each voxel in a local neighbourhood and assign weights accordingly [16].

The Probabilistic Patch-Based Model (PPBM) was firstly proposed by Sabuncu *et al.* [243] and then improved by Bai *et al.* [25]. The improved model accounts for the potential registration error using the following strategy: for each voxel at $\mathscr{S}_i$ in the target image, the label depends not only on the corresponding i^{th} voxels in the atlases but also a set of candidates in a local neighbourhood $\mathscr{M}_i$ with $i \in \mathscr{M}_i$ in each atlas.

The PPBM consists of two components: the intensity likelihood and the label likelihood. The intensity likelihood models the intensity difference between the target and candidate voxels as a Gaussian distribution. For each voxel at $\mathscr{S}_i$ in the target image y, given the candidate location $j \in \mathscr{M}_i$ in the atlas a, the probability of y_i is computed as

$$P(y_i|j,a) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{D(B_i,B_j^a)}{2\sigma_1^2}\right], \tag{6.17}$$

where σ_1 denotes the standard deviation of the Gaussian distribution. The intensity difference between the two voxels is computed by the difference between their corresponding patches B_i and B_j^a as $D(B_i,B_j^a)$, the mean squared difference between the two patches. The label likelihood models the registration error as a Gaussian distribution with the standard deviation σ_2. Similarly, the probability of the voxel at $\mathscr{S}_i$ with the label l is computed as

$$P(x_i = l|j,x^a) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{d^2(i,j)}{2\sigma_2^2}\right] \xi_{x_j^a,l}, \tag{6.18}$$

where x_j^a is the label of the candidate and $d(i,j)$ is the Euclidean distance between the target voxel and the candidate.

Taking all the candidates in all the atlases into account, the optimal label of the voxel at $\mathscr{S}_i$ is determined by

$$x_i^* = \arg\max_{l\in\mathscr{L}} \sum_{a\in\mathscr{A}} \sum_{j\in\mathscr{M}_i} P(y_i|j,a)P(x_i = l|j,x^a). \tag{6.19}$$

Combination of Global and Local Label Fusion

It is straightforward to assign a global weight to each atlas in the local label fusion strategies. By combining MI with PPBM (MI-PPBM), Eq. 6.19 is reformulated as

$$x_i^* = \arg\max_{l\in\mathscr{L}} \sum_{a\in\mathscr{A}} \sum_{j\in\mathscr{M}_i} \frac{I(a;y)}{I_{max}} P(y_i|j,a)P(x_i = l|j,x^a). \tag{6.20}$$

6.2.4 Combination of Image Intensity, Local and Multi-Atlas Priors: KM-MRF-MAS

The multi-atlas prior can easily be combined with KM-MRF by reformulating the energy functional from Eq. 6.2 as

$$\begin{aligned} x^* &= \arg\min_{x\in\mathscr{L}^N}(-\ln P(y|x)-\ln P(x)) \\ &= \arg\min_{x\in\mathscr{L}^N}(-\ln P(y|x)-\ln(P_{Lpr}(x)P_{Apr}(x))) \\ &\propto \arg\min_{x\in\mathscr{L}^N}(E_{in}(x)+E_{Lpr}(x)+\gamma E_{Apr}(x)). \end{aligned} \tag{6.21}$$

The probability of the prior classification $P(x)$ is determined by both the local and multi-atlas prior probabilities which are represented as P_{Lpr} and P_{Apr}, respectively. Then the two prior probabilities are rewritten as energy terms $E_{Lpr}(x)$ and $E_{Apr}(x)$. The effect of the local prior is balanced with that of the image intensity by the spatial parameter β embodied in $E_{Lpr}(x)$. γ is introduced to balance the effect of the multi-atlas prior with that of the intensity and neighbouring information. The local prior energy $E_{Lpr}(x)$ is defined by Eq. 6.10 and the multi-atlas prior energy $E_{Apr}(x)$ which is based on each label fusion strategy, is defined in Appendix A. When combining the three energy terms for minimisation, all the constants can be ignored. The classification approach is almost the same as Algorithm 2 except that the energy functional in step 2 is replaced by Eq. 6.21.

6.3 Experiments and Results

6.3.1 Databases

We used one synthetic database BrainWeb and three real databases IBSR18, IBSR20 and MICCAI2013 described in Section 2.4. For BrainWeb, 18 T1-weighted brain MR images with various levels of noise (0%, 1%, 3%, 5%, 7% and 9%) and intensity inhomogeneity (0%, 20% and 40%) were used. The image resolution is $1mm \times 1mm \times 1mm$. The segmentation of a normal anatomical model provided in the database was used to perform the skull stripping and generate the ground truth for the test data. The brain mask was obtained from combining the segments of WM, GM and CSF and then applied to all the simulated images. The ground truth of the three tissues was also produced from their fuzzy models by assigning each voxel the dominant tissue class. For IBSR18 and IBSR20, 18 and 20 normal T1-weighted brain images with their

manual segmentations were used for the experiments, respectively. The brain mask of each volume was generated by filling the holes of the combination of the manually segmented brain tissues. The filled voxels were then combined into the ground truth of CSF. For MICCAI2013, twenty 3T scans are available including 5 cases for training and 15 for testing. We only used T1-weighted scans with volume size $240 \times 240 \times 48$ and resolution $0.958mm \times 0.958mm \times 3mm$.

As mentioned above, for BrainWeb and IBSR databases, the ground truth of brain segmentation was used for all the images. We did not apply any other skull stripping step in order to make sure the tissue classification performance is not affected by the skull stripping methods. For MICCAI2013, because the ground truth for the test images were not provided, the MAS method was applied to remove the skulls from the whole-head images. The five training images were considered as atlases. The brain mask of each training image was generated from the label map of tissue segmentation. The voxels of three tissues (WM, GM and CSF) were considered as foreground and the other voxels were considered as background. By registering each training image to the test image, 5 candidate brain masks could be obtained for each test image. These candidate brain masks were combined by applying majority voting to generate the final brain mask for the test images. Image intensity normalisation was performed for all four databases in the process of atlas selection because it is necessary for image registration. Other preprocessing steps, such as intensity inhomogeneity or noise removal, were not performed.

6.3.2 Evaluation Measures

For each tissue class, we calculate the Dice similarity coefficient (DSC) [87] defined in Section 3.2.1 to measure the spatial overlap accuracy between the classification result and the ground truth. The DSC is calculated for each tissue class: DSC_{CSF} for CSF, DSC_{GM} for GM and DSC_{WM} for WM, respectively. The overall accuracy for classifying the three tissues is defined by

$$AC = \frac{TP_{WM} + TP_{GM} + TP_{CSF}}{|S|}, \tag{6.22}$$

where $|S|$ denotes the number of voxels in the brain mask S; TP_{CSF}, TP_{GM} and TP_{WM} represent the numbers of voxels correctly classified as CSF, GM and WM, respectively. AC calculates the proportion of all the correctly classified voxels within the brain mask. The values of both metrics are in the range [0, 1]. In our experiments, both metrics are

used to measure the classification on BrainWeb and IBSR databases. For MRBrainS13, only the evaluation results measured by DSC are provided by the challenge.

In our experiments, a paired-sample t-test [15] is adopted to analyse the impacts of the parameter settings on the performance of specific methods. It measures the difference between two sets of classification results with the significance level set at 0.05. When comparing the performance of different methods with regard to several metrics on the same database, repeated-measures MANOVA [189] is applied. Both of them are described in more detail in Section 3.5.

6.3.3 Parameters Optimisation

For BrainWeb and IBSR databases, we performed k-fold cross validation to optimise the parameters for each method. The choice of k depends on the number of brain volumes in the database. We used 3-fold cross validation for BrainWeb and IBSR18 (both contain 18 volumes so 6 per fold), while 4-fold cross validation was used for IBSR20 (containing 20 volumes so 5 per fold). For MICCAI2013, the five training images were used to optimise the parameters involved in all the methods.

For KM, we set the number of clusters K as 3 because we aim to segment the brain tissue into three classes. For KM-MRF, a range of values were tested for the parameters β and the size of the local neighbourhood $|\mathscr{N}|$. At first, we set an initial value for β and evaluated the performance of KM-MRF when varying $|\mathscr{N}|$ between 6 and 26 (as shown in Fig. 6.4). After applying k-fold cross validation, the optimal $|\mathscr{N}|$ was obtained. Then we set $|\mathscr{N}|$ to the optimal value and tuned β in a range around its initial value, for which we again use k-fold cross validation. Finally, for each database, the brain tissue classification results were produced from applying KM-MRF with the corresponding optimised parameters. For MAS, we optimised the number of atlases fused n_{atlas}. In addition, 4 parameters involved in PPBM and MI-PPBM, the size of the local neighbourhood $|\mathscr{M}|$, the patch size $|B|$, σ_1 and σ_2, were also optimised. A similar approach was used as for KM-MRF: when optimising one parameter, the others were set to the initial or optimised values. A range of values were tested in optimising each parameter. For KM-MRF-MAS, we used the parameters optimised for KM-MRF and MAS. The additional parameter γ was optimised by evaluating the performance of KM-MRF-MAS with varying γ around its initial value.

Table 6.1 Optimised parameters for all the methods applied on real data

Method		KM-MRF		MAS					KM-MRF-MAS
Parameter		β	$\lvert\mathcal{N}\rvert$	n_{atlas}	$\lvert B\rvert$	$\lvert\mathcal{M}\rvert$	σ_1	σ_2	γ
Database	BrainWeb	[1, 1.5]	6	-	-	-	-	-	-
	IBSR18	[100, $+\infty$)	10	5, 6	11	19	0.5	[1, 1.5]	[400, $+\infty$)
	IBSR20	[20, $+\infty$)	10	3, 5, 7	11	19	[20, $+\infty$)	10	[400, $+\infty$)
	MICCAI2013	0.3	6	3	7	7	0.1	2	0.3

- β is the spatial parameter used to balance the effect of the image intensity and the local prior; $|\mathcal{N}|$ represents the size of the local neighbourhood in KM-MRF; n_{atlas} stands for the number of atlases fused in MAS; $|B|$ and $|\mathcal{M}|$ represent the patch size and the size of the local neighbourhood considered in MAS using PPBM for label fusion; σ_1 and σ_2 are also the parameters of PPBM; γ is used to balance the effect of the multi-atlas prior with that of the image intensity/local prior.
- '[ ]' or '[ )' is used to represent the range of the parameter values.
- '-' means the method is not applied on the corresponding database.

In the optimisation process, different initial values for β (e.g. 1, 10, 100) were tested and we chose the range which produces the best segmentation results as the optimal parameter. The validation was repeated three times in optimising each parameter. It was found that the standard deviations among the segmentation results generated from repeating the optimisation process are very small.

In Table 6.1, we list the optimal values for all the parameters involved in all the methods applied on various databases.

6.3.4 Validation of KM-MRF

Experiments on simulated data

Firstly, we tested the model KM-MRF described in Section 6.2.2 on the 18 simulated brain images in BrainWeb. The spatial parameter β was set in the range [1, 1.5] and 6 nearest neighbours (see Fig. 6.4(a)) were taken into account at each voxel. For each brain image with specific levels of noise and intensity homogeneity, the classification accuracy *AC* was compared between applying KM-MRF and KM which uses the image intensity only. As illustrated in Fig. 6.1, when the noise level is lower than 5%, KM slightly outperforms KM-MRF; especially when both the noise and the bias field levels are very low (e.g. the images with 0% noise 0% bias field and 1% noise 0% bias field), KM-MRF performs much worse than KM. However, when the noise level increases (5% or higher), the classification benefits greatly from the neighbouring information. For most images with a certain level of noise, both methods deteriorate from increasing the bias field level; for most images with a certain level of bias field, both methods also perform worse when adding more noise but the *AC* of applying KM declines more steeply than when using KM-MRF especially when the noise level reaches 5%.

A qualitative comparison is given in Fig. 6.2. For the image with low noise levels , KM slightly outperforms KM-MRF and we can hardly observe any difference between their classification results based on a single slice (see b and c). However, for higher noise levels d, KM-MRF distinctly improves the performance with the local prior taken into account (see f and g).

Experiments on real data

We also tested KM-MRF on three real databases. The values of the parameters β and $|\mathcal{N}|$ applied on each database are listed in Table 6.1. β was set to 100 and 20 for IBSR18 and IBSR20, respectively. The impacts of $|\mathcal{N}|$ and β on the classification

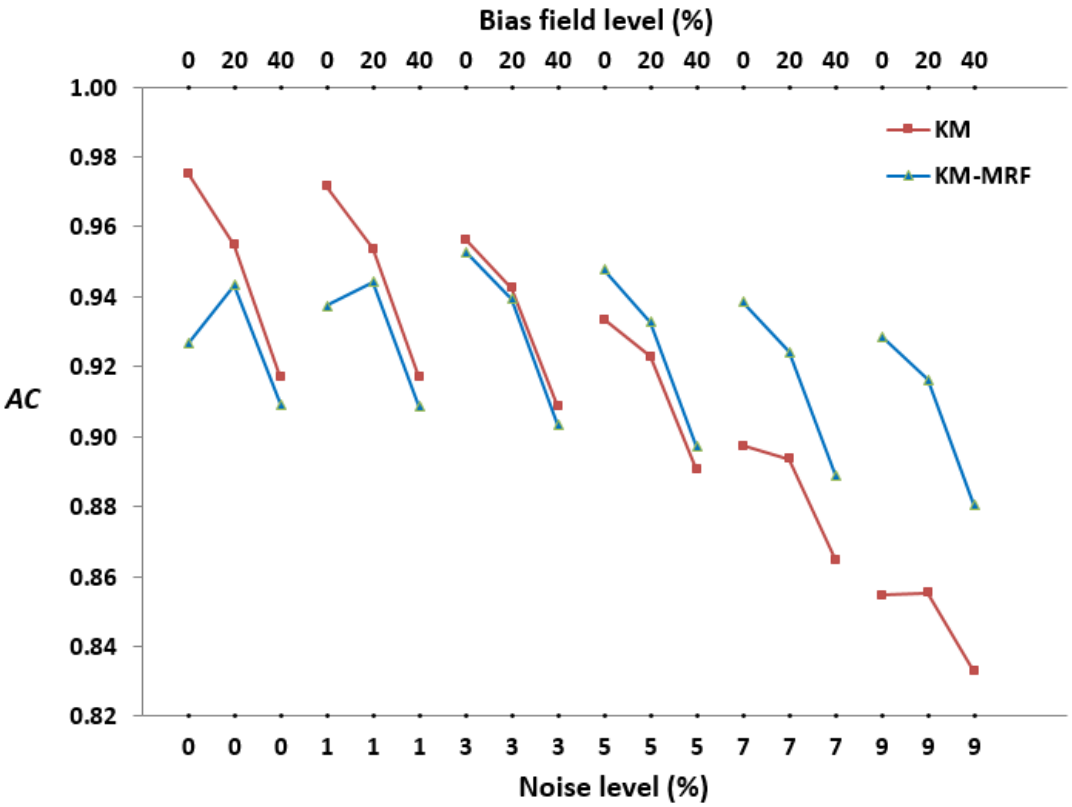
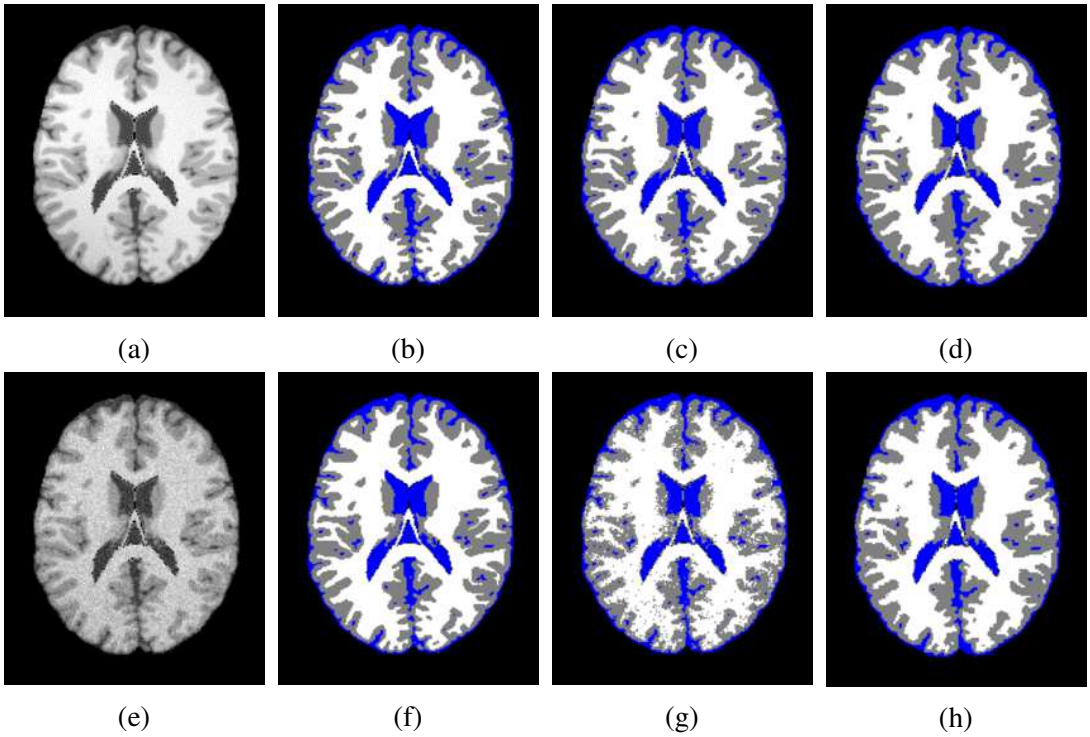


Fig. 6.1 Comparison of the *AC* (see Eq. 6.22) values between applying KM-MRF and KM on the brain images with various levels of noise and bias field in BrainWeb.

accuracy will be analysed later. The classification results are compared with those produced by KM. The repeated-measures MANOVA is applied to test the differences between the performance of two methods and the results are listed in Table 6.8. The multivariate outcome indicates that significant differences exist between the performance of two methods with respect to the combination of all metrics. The univariate outcomes demonstrate that KM-MRF performs significantly better than KM in terms of the overall accuracy (*AC*) and the classification accuracies of CSF (DSC_{CSF}) and GM (DSC_{GM}). Consistent results are obtained from the qualitative comparisons illustrated in Fig. 6.3 where the difference maps between the classification results generated by each method and the ground truth are displayed for each tissue class. We can see that the label map generated by KM-MRF is more contiguous than that produced by KM because the local prior encourages each tissue class to be more continuous; the false positives are decreased in CSF; a number of false negatives are eliminated in GM. The results in Table 6.8 also show that KM-MRF performs worse (non-significantly on IBSR18 and significantly on IBSR20) than KM in the classification of WM; however, taking the classification of all three tissues into



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 6.2 Axial slices of two volumes with 1% noise and 40% bias field, d with 7% noise and 20% bias field; their ground truth a and e; the classification results of applying KM (b and f) and KM-MRF (c and g). Three tissues (CSF, GM and WM) in the ground truth and classifications are labelled in blue, grey and white, respectively. The overall accuracy *AC* and the accuracies of classifying three tissues DSC_{CSF}, DSC_{GM} and DSC_{WM} measured for each classified volume are: 0.92, 0.94, 0.91 and 0.91 for b; 0.91, 0.93, 0.90 and 0.90 for c; 0.89, 0.92, 0.88 and 0.89 for f; 0.93, 0.93, 0.92 and 0.93 for g.

account, the overall classification accuracy of KM-MRF is significantly higher than KM. For MICCAI2013, KM-MRF performs significantly better than KM in terms of the classification accuracies of all three tissue types. Based on these comparisons, the conclusion is drawn that the classification is significantly improved by taking the local prior into account on real databases.

Impact of the size of the local neighbourhood

When β is initialised as 10, the number of neighbours for each voxel in the test image of IBSR20 is varied between 6 and 26 as illustrated in Fig. 6.4. The impact of the local neighbourhood selection on the overall classification accuracy using KM-MRF is displayed in Fig. 6.5. The paired-sample t-test is used to test the difference between two sets of classification results produced with different number of neighbours. The *AC* values generated by varying the size of the local neighbourhood and the p-values

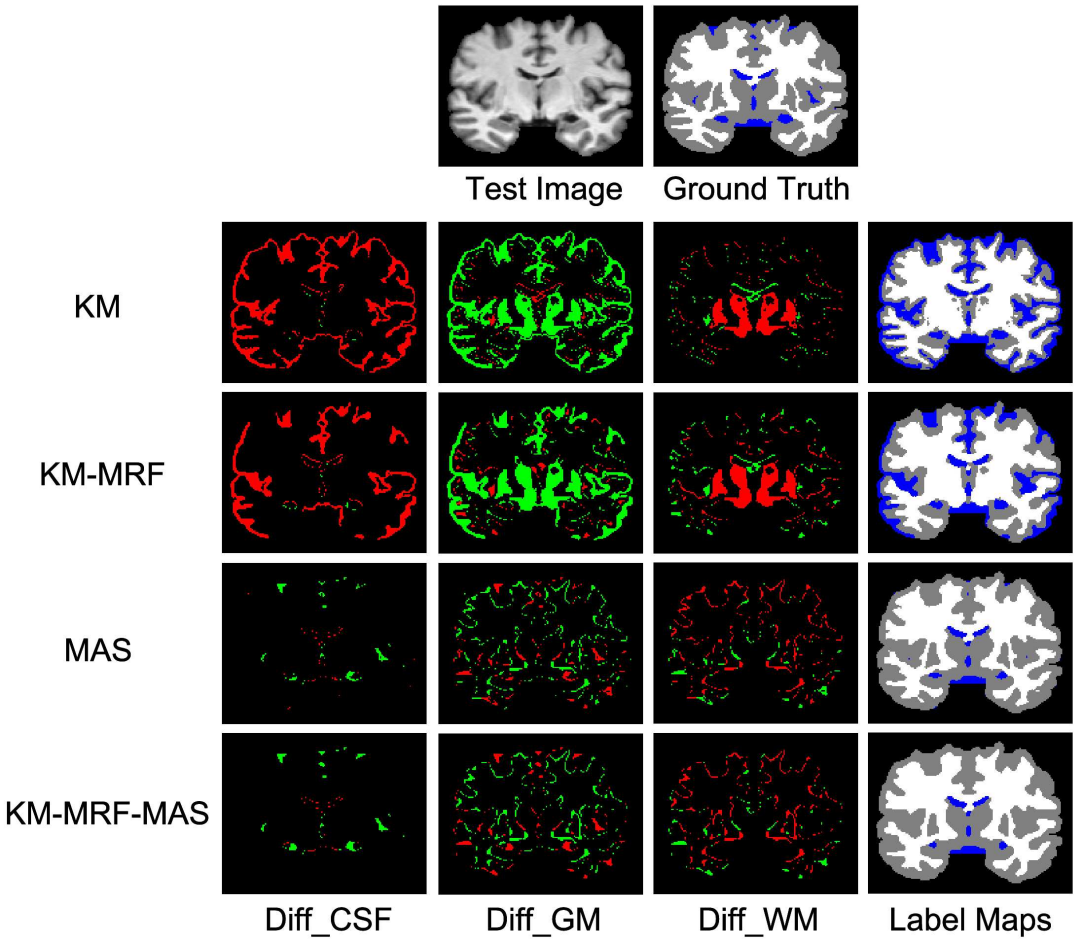


Fig. 6.3 Top row: coronal slice of a representative volume in IBSR18 and its ground truth of tissue classification. The difference maps of three tissues (Diff_CSF, Diff_GM and Diff_WM) and the label maps are produced by KM, KM-MRF, MAS and KM-MRF-MAS, respectively. Three tissues (CSF, GM and WM) in the ground truth and classifications are labelled in blue, grey and white colours, respectively. The red and green points in the difference maps represent the false positives and false negatives, respectively. The classification accuracies of the volumes produced by the four methods KM, KM-MRF, MAS and KM-MRF-MAS are: 0.71, 0.75, 0.92 and 0.92 measured by AC; 0.18, 0.24, 0.74 and 0.74 measured by DSC_{CSF}; 0.69, 0.75, 0.94 and 0.93 measured by DSC_{GM}; 0.90, 0.88, 0.92 and 0.91 measured by DSC_{WM}.

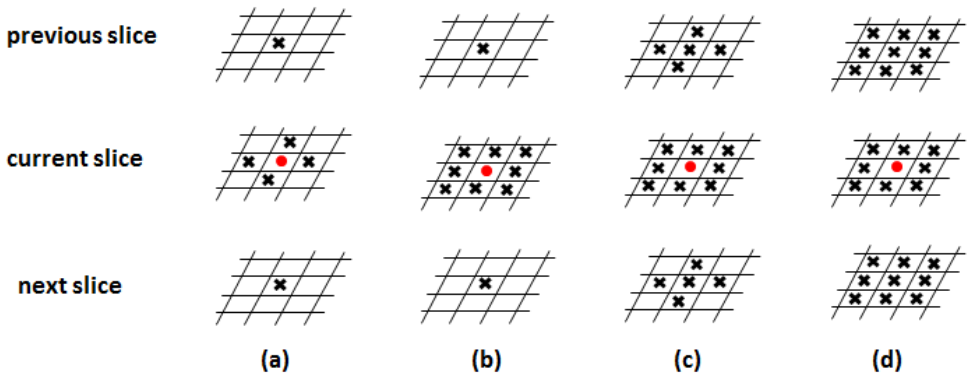


Fig. 6.4 Selections of the local neighbourhood for each voxel. '•' represents the current voxel and '×' represents its selected neighbour. The neighbours are not always symmetric around the central voxel because the image resolutions vary along three dimensions.

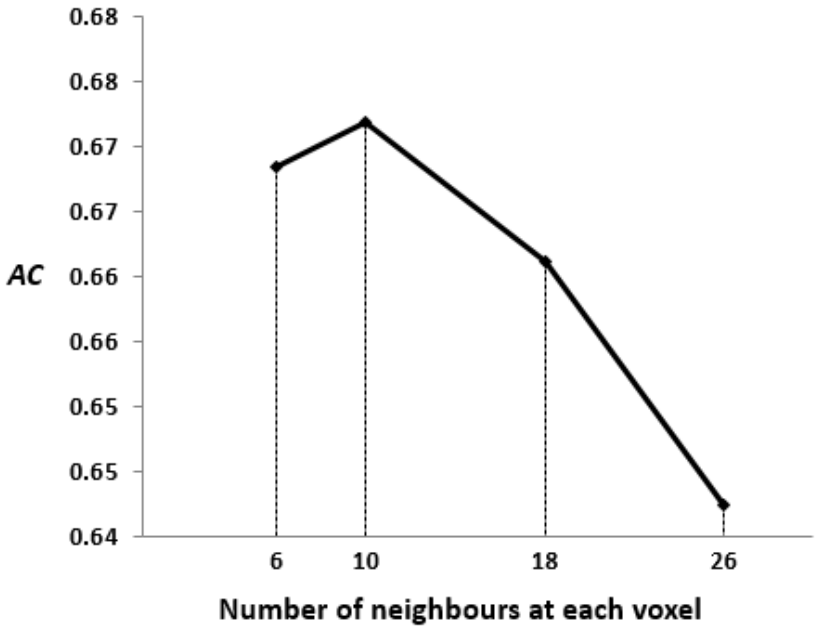


Fig. 6.5 *AC* values averaged over IBSR20 when varying the number of neighbours considered at each voxel. The standard deviations of the results obtained across different number of neighbours are comparable (between ± 0.15 and ± 0.18). The standard deviations are large compared with the mean values because of outliers.

tested between any two sets of results produced with different numbers of neighbours are listed in Table 6.2. At first, *AC* rises when the size of the local neighbourhood increases from 6 to 10. And it declines continuously when the number of neighbours is further increased to 18 and 26. Similar analyses were conducted on IBSR18 and MICCAI2013.

Table 6.2 Impact of the size of the local neighbourhood on the performance of KM-MRF applied on IBSR20 and measured by *AC*

	6	10	18	26
6	$\mathbf{0.67 \pm 0.18}$	0.313	0.279	$0.008^{\star}$
10	–	$\mathbf{0.67 \pm 0.17}$	$0.009^{\star}$	$0.001^{\star}$
18	–	–	$\mathbf{0.66 \pm 0.16}$	$< 0.001^{\star}$
26	–	–	–	$\mathbf{0.64 \pm 0.15}$

- The numbers in the first column and the first row are the numbers of neighbours at each voxel tested for KM-MRF on IBSR20. The values highlighted in bold show the mean *AC* values and the standard deviations produced by KM-MRF with the corresponding size of the local neighbourhood. Each of the other numbers in the table shows the *p*-value tested between any two sets of classification results produced with different numbers of neighbours in the corresponding row and column. The significant differences are indicated by $\star$ at the top right of *p*-values.
- The *p*-values in the table are symmetric, so we only show half of them and the omitted ones are replaced by '–'.

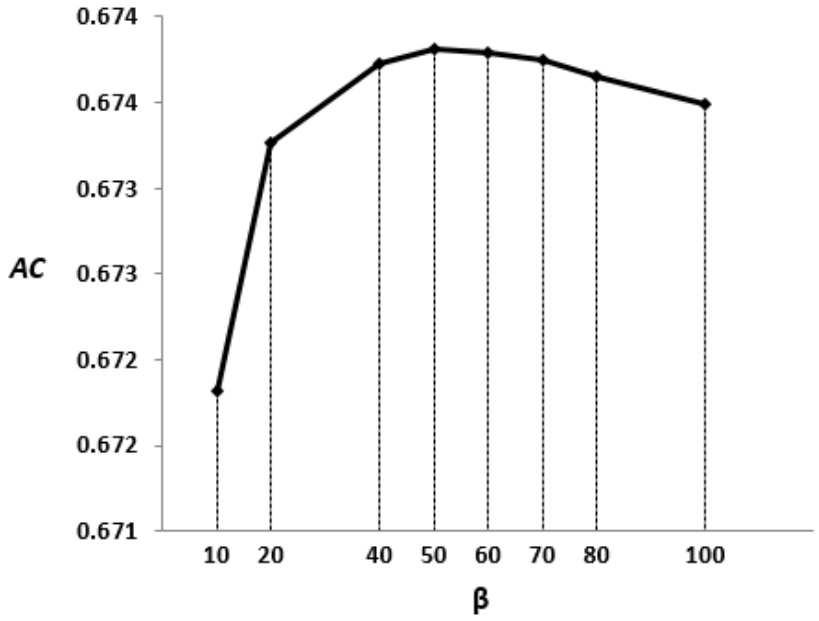


Fig. 6.6 *AC* values averaged over IBSR20 when varying the parameter β. The standard deviations of the results obtained across different β values are all ± 0.17. The standard deviations are large compared with the mean values because of the outliers.

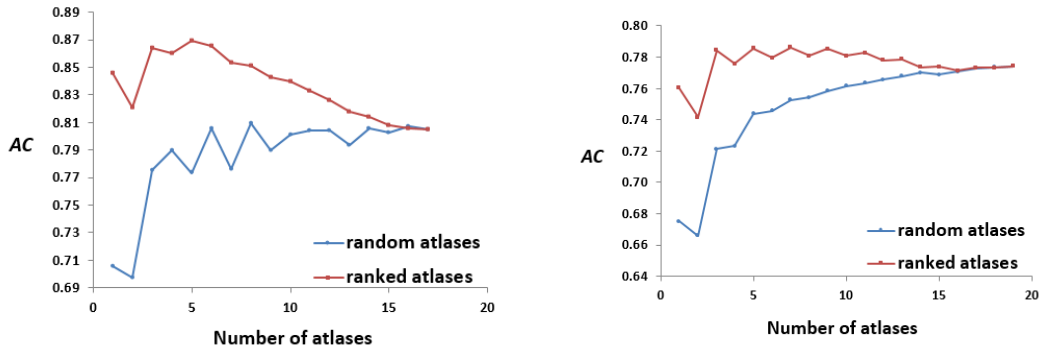
Impact of β

The impact of the spatial parameter β on the overall classification accuracy of KM-MRF is analysed in Fig. 6.6. Ten neighbours (see Fig. 6.4(b)) are considered for each voxel and the model is applied on the brain images in IBSR20 with a range of β values. The paired-sample t-test was used to test the difference between any two sets of classification results produced with different β values. The *AC* value generated by each β and the p-values tested between any two sets of results produced with different β values are listed in Table 6.3. β increases from 10. During this process, *AC* increases and becomes stable when β reaches 20. After that, no significant difference was found when further increasing β compared with the results obtained when β is 20. We can conclude that KM-MRF attains a stable performance when the parameter β is set beyond a threshold. An analysis with similar results was conducted on IBSR18.

Table 6.3 Impact of β on the performance of KM-MRF applied on IBSR20 and measured by *AC*

	10	20	40	60	80	100
10	$\mathbf{0.67 \pm 0.17}$	$0.001^{\star}$	$0.003^{\star}$	$0.002^{\star}$	$0.002^{\star}$	$0.006^{\star}$
20	–	$\mathbf{0.67 \pm 0.17}$	$0.101^{\star}$	0.055	0.147	0.416
40	–	–	$\mathbf{0.67 \pm 0.17}$	0.318	0.457	0.201
60	–	–	–	$\mathbf{0.67 \pm 0.17}$	$0.017^{\star}$	0.060
80	–	–	–	–	$\mathbf{0.67 \pm 0.17}$	0.222
100	–	–	–	–	–	$\mathbf{0.67 \pm 0.17}$

- The numbers in the first column and the first row are the β values tested for KM-MRF on IBSR20. The values highlighted in bold show the mean *AC* values and the standard deviations produced by KM-MRF with the corresponding β. Each of the other numbers in the table shows the *p*-value tested between two sets of classification results produced with a pair of β values in the corresponding row and column. The significant differences are indicated by $\star$ at the top right of *p*-values.
- The *p*-values in the table are symmetric, so we only show half of them and the omitted ones are replaced by '–'.



(a) (b)

Fig. 6.7 *AC* values of applying MAS by increasing the number of ranked and random atlases on IBSR18 (a) and IBSR20 (b). MV is used for multi-atlas label fusion.

6.3.5 MAS Based on Various Label Fusion Strategies

The performance of MAS based on various label fusion strategies was validated on the real databases IBSR18 and IBSR20. For every brain image, the other images in the same database together with their label maps were considered as the atlases. All the atlas images were nonrigidly registered to the target image. Subsequently the similarity between each pair of the registered image and the target was calculated based on the mutual information and the atlases were ranked according to the similarities. The top ranked atlases were selected and their label maps were propagated to the target image to perform the tissue classification. The validation was not conducted on BrainWeb since all the simulated images in this database have the same ground truth.

Impact of the atlas selection

To investigate the necessity of the atlas selection, we tested MAS by increasing the number of ranked and randomly selected atlases on both IBSR18 and IBSR20. MV is utilised for multi-atlas label fusion due to its simplicity. The classification accuracy is measured by *AC* and averaged over the database as shown in Fig. 6.7. For the two approaches using the same number of atlases, the standard deviation produced by fusing the ranked atlases is always smaller than that produced by fusing the randomly selected atlases. The paired-sample t-test is used to test the difference between any two sets of classification results produced by approaches using the same number of ranked and random atlases as shown in Table 6.4.

From Fig. 6.7 and Table 6.4, it is observed that at first MAS using the ranked atlases significantly outperforms that using the same number of random atlases on both databases. When increasing the number of atlases fused, the difference between

Table 6.4 Comparison of performance of MAS using random and ranked atlases

Number of atlases	IBSR18	IBSR20
1	$3.84e-04^{\star}$	$9.31e-05^{\star}$
2	$7.36e-04^{\star}$	$7.41e-05^{\star}$
3	$3.60e-05^{\star}$	$1.20e-05^{\star}$
4	$3.71e-04^{\star}$	$4.54e-07^{\star}$
5	$3.36e-04^{\star}$	$1.18e-05^{\star}$
6	$0.002^{\star}$	$0.001^{\star}$
7	$2.17e-04^{\star}$	$1.41e-06^{\star}$
8	$1.94e-04^{\star}$	$2.39e-05^{\star}$
9	$1.66e-04^{\star}$	$5.94e-05^{\star}$
10	$0.002^{\star}$	$9.61e-05^{\star}$
11	$3.19e-04^{\star}$	$1.77e-05^{\star}$
12	$0.006^{\star}$	$9.06e-04^{\star}$
13	$0.007^{\star}$	$4.48e-04^{\star}$
14	0.075	0.411
15	0.189	0.164
16	0.677	0.720
17	–	0.794
18	–	0.868

- Each value in the table represents the p-value between the performance of MAS using the number of random and ranked atlases indicated in its corresponding row. Significant difference is indicated by the $\star$ at the top right of each p-value. The comparison is conducted on both IBSR18 and IBSR20. The classification performance is measured by *AC*.

Table 6.5 Analysis of the impact of atlas selection

Database	Approach	n_{atlas}	AC	p-value
IBSR18	Ranked	5, 6	0.87 ± 0.06	< 0.001
	Random	6, 8, 11	0.79 ± 0.08	
IBSR20	Ranked	3, 5, 7	0.78 ± 0.05	0.13
	Random	18, 19	0.77 ± 0.08	

- n_{atlas} represents the optimal number of atlases; MV is used for label fusion.

the AC values obtained from fusing the ranked and random atlases decreases until it becomes nonsignificant (when fusing 14 atlases or more) and finally vanishes (when all the atlases are fused). Thus it is concluded that the atlases contribute unequally to the classification of brain images and the atlas selection approach picks out the most effective ones.

Then we run k-fold cross validation to optimise the number of ranked and random atlases fused in MAS on each database. In Table 6.5, we listed the comparisons of the two approaches applied on the two databases. The tissue classification performance was measured by AC, which is the mean value across the whole database in three validations. The paired-sample t-test was used to test the difference between two sets of AC values generated from two approaches. It is shown that for IBSR18, applying atlas selection not only decreases the number of atlases fused but also significantly improves the segmentation performance. For IBSR20, no significant difference is found between the two sets of classification results measured by AC. However, the number of atlases fused is substantially reduced by applying atlas selection.

Therefore from a performance point of view, applying atlas selection significantly improves the classification accuracy or decreases the number of atlases to be fused without deteriorating the result. Reducing the number of atlases involved in MAS reduces computation for label fusion; nevertheless, the great amount of registration required for atlas selection still makes it time-consuming. This is the main weakness of adopting the atlas selection in MAS.

Table 6.6 Comparisons of the performance of MAS using different label fusion strategies

Database	Method	*AC*	DSC_{CSF}	DSC_{GM}	DSC_{WM}
IBSR18	MV	0.87 ± 0.06	0.56 ± 0.13	0.90 ± 0.04	0.83 ± 0.14
	MI*	$0.87 \pm 0.06^{\star}$	$0.58 \pm 0.12^{\star}$	$0.90 \pm 0.04^{\star}$	0.83 ± 0.14
	PPBM**	$0.88 \pm 0.06^{\star\star}$	$0.70 \pm 0.13^{\star\star}$	$0.91 \pm 0.03^{\star\star}$	0.83 ± 0.14
	MI-PPBM**	$0.89 \pm 0.06^{\star\star}$	$0.70 \pm 0.13^{\star\star}$	$0.92 \pm 0.03^{\star\star}$	0.83 ± 0.14
IBSR20	MV	0.78 ± 0.05	0.39 ± 0.08	0.83 ± 0.04	0.76 ± 0.06
	MI*	$0.79 \pm 0.04^{\star}$	$0.41 \pm 0.08^{\star}$	$0.83 \pm 0.03^{\star}$	0.76 ± 0.06
	PPBM**	$0.82 \pm 0.04^{\star\star}$	$0.57 \pm 0.09^{\star\star}$	$0.86 \pm 0.03^{\star\star}$	$0.76 \pm 0.06^{\star}$
	MI-PPBM**	$0.82 \pm 0.04^{\star\star}$	$0.57 \pm 0.09^{\star\star}$	$0.86 \pm 0.03^{\star\star}$	$0.76 \pm 0.05^{\star}$

- Each value represents the mean accuracy and the standard deviation of applying MAS based on each label fusion strategy on the database.
- The $\star$s at the top right of each value indicate the significant differences of the corresponding result compared with all the others derived by applying other label fusion methods on the same database and measured by the same metric. The $\star$s at the top right of each method indicate the significant differences of the performance of the corresponding method compared with all the other methods. The result or method labelled with more $\star$s is significantly better than those with fewer or no $\star$s and the results or methods labelled with equal number of $\star$s have no significant difference between each other.

Impact of the label fusion strategy

Four label fusion strategies described in Section 6.2.3 were considered for MAS and the classification results were compared on both IBSR18 and IBSR20. Five and three top ranked atlases were fused in MAS applied on IBSR18 and IBSR20, respectively. The other parameter settings of PPBM on IBSR18 and IBSR20 are listed in Table 6.1. σ_1 was set to 20 when applying PPBM on IBSR20; σ_2 was set to 1 when applying PPBM on IBSR18. The parameters of MI-PPBM were the same as those of PPBM on each database.

The repeated-measures MANOVA is applied to test the performance differences of MAS using four label fusion strategies and the results are listed in Table 6.6. Significant differences are detected among the four methods with respect to the combination of four metrics on both databases. For IBSR18, the performance of four strategies is significantly different in terms of the overall accuracy (*AC*) and the classification accuracies of CSF (DSC_{CSF}) and GM (DSC_{GM}). The *post hoc* pairwise comparison results of the four methods are consistent with respect to these three metrics: MI significantly outperforms MV; PPBM and MI-PPBM perform equally and they perform better than both MV and MI. No significant differences exist among the four strategies in the classification of WM. For IBSR20, significant differences exist among the performance of four methods in terms of each metric. The pairwise comparison results with regard to *AC*, DSC_{CSF} and DSC_{GM} are the same with those for IBSR18. In the classification of WM, PPBM performs equally with MI-PPBM and they perform better than both MV and MI. Based on the comparison results on IBSR20, the performance of the four methods is ranked exactly the same as on IBSR18: MI outperforms MV; PPBM outperforms both MV and MI; MI-PPBM performs equally to PPBM.

We can draw the conclusion that in MAS, the mutual information based global weighting scheme MI performs better than MV; the probabilistic local label fusion model PPBM achieves better results than global models; finally the combination of global and local label fusion strategies MI-PPBM performs equally to PPBM.

6.3.6 Comparisons of Classification Models Using Different Image Based Features

Impact of γ on KM-MRF-MAS

The impact of the parameter γ on the overall classification accuracy of applying KM-MRF-MAS on IBSR20 is analysed in Fig. 6.8. In this experiment, γ is varied from 10 to 1000. The paired-sample t-test is used to test the difference between two sets

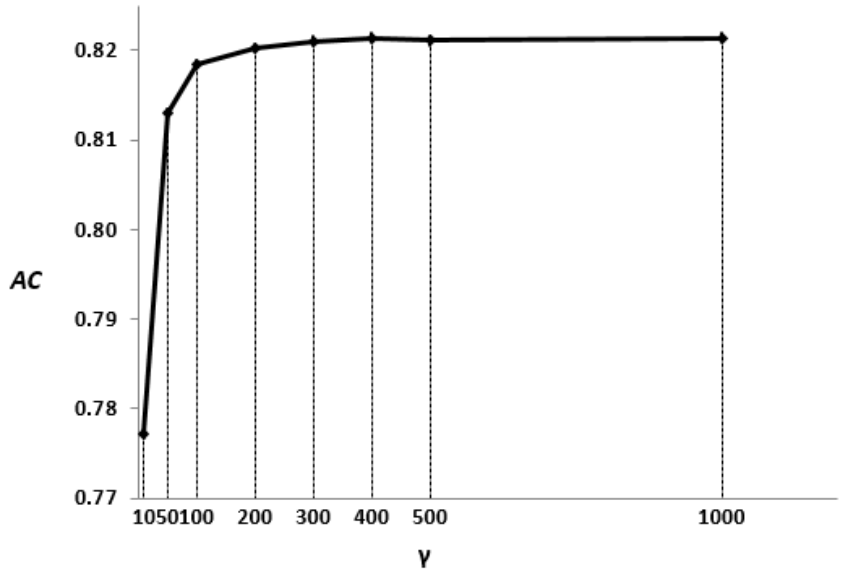


Fig. 6.8 Mean *AC* values of applying KM-MRF-MAS on IBSR20 when varying the parameter γ. The standard deviation is ± 0.08 when γ is 10 and ± 0.04 when γ increases from 50 to 1000. PPBM is used for the multi-atlas label fusion.

of classification results produced with different γ values. The *AC* values generated by varying γ and the p-values tested between any two sets of results produced with a pair of γ values are listed in Table 6.7. The classification accuracy *AC* is significantly improved when γ is increased from 10 to 400. No significant improvement is obtained when further increasing γ. It is therefore concluded that the performance of KM-MRF-MAS is not affected by the parameter γ when it is set beyond a threshold. In our experiments, we set γ to 400 for both IBSR20 and IBSR18 (the latter based on a similar analysis).

Table 6.7 Impact of γ on the performance of KM-MRF-MAS applied on IBSR20 and measured by *AC*

	10	100	200	300	400	500	1000
10	$\mathbf{0.78 \pm 0.08}$	$< 0.001^{\star}$	$< 0.001^{\star}$	$< 0.001^{\star}$	$< 0.001^{\star}$	$< 0.001^{\star}$	$< 0.001^{\star}$
100	–	$\mathbf{0.82 \pm 0.04}$	$0.002^{\star}$	$< 0.001^{\star}$	$< 0.001^{\star}$	$0.001^{\star}$	$< 0.001^{\star}$
200	–	–	$\mathbf{0.82 \pm 0.04}$	$< 0.001^{\star}$	$< 0.001^{\star}$	$0.003^{\star}$	$0.004^{\star}$
300	–	–	–	$\mathbf{0.82 \pm 0.04}$	$< 0.001^{\star}$	0.744	0.138
400	–	–	–	–	$\mathbf{0.82 \pm 0.04}$	$< 0.001^{\star}$	0.937
500	–	–	–	–	–	$\mathbf{0.82 \pm 0.04}$	$0.034^{\star}$
1000	–	–	–	–	–	–	$\mathbf{0.82 \pm 0.04}$

- The numbers in the first column and the first row are the γ values tested for KM-MRF-MAS on IBSR20. The values highlighted in bold show the mean *AC* values and the standard deviations produced by KM-MRF-MAS with the corresponding γ. Each of the other numbers in the table shows the *p*-value tested between two sets of classification results produced with a pair of γ values in the corresponding row and column. The significant differences are indicated by $\star$ at the top right of *p*-values.
- The *p*-values in the table are symmetric, so we only show half of them and the omitted ones are replaced by '–'.

Comparisons of different classification models

Based on the conclusions we drew above, PPBM was adopted in the combined KM-MRF-MAS model. We compared the performance of the four tissue classification models on the three real databases. The optimal values were used for the parameters involved in each method applied on each database as listed in Table 6.1. The repeated-measures MANOVA was applied to test the performance differences among the four models on the three databases and the results are listed in Table 6.8. The multivariate outcome tells that significant differences exist among the four methods with respect to the combination of the four metrics on all three databases. For IBSR18, the univariate outcomes demonstrate that significant differences exist among the four methods in terms of the overall accuracy (AC) and the classification accuracies of CSF (DSC_{CSF}) and GM (DSC_{GM}). The four models perform similarly in the classification of WM. The *post hoc* pairwise comparisons of the four models are conducted with respect to the metrics AC, DSC_{CSF} and DSC_{GM}. For IBSR20, significant differences exist among the four methods with respect to all four metrics. The results of *post hoc* pairwise comparisons of the four models regarding AC, DSC_{CSF} and DSC_{GM} are the same with those for IBSR18. For the classification accuracy of WM (DSC_{WM}), it shows that KM-MRF performs significantly worse than the other methods.

By comparing the performance of the three methods KM, KM-MRF and KM-MRF-MAS applied on IBSR databases, it is observed that introducing additional information including the local and multi-atlas priors significantly improves the tissue classification performance with respect to the overall accuracy and the classification accuracies of CSF and GM. Fig. 6.3 gives an example of the qualitative comparisons on a typical volume of IBSR18, which are consistent with the quantitative comparisons. Comparing the difference maps produced by KM, KM-MRF and KM-MRF-MAS, it is observed that the total numbers of false positives and false negatives are dramatically decreased in the classification of CSF and GM by introducing additional information. From the label map generated by each method, it can be seen that acceptable tissue classification is achieved by KM based on the image intensity prior, which provides an initialisation for KM-MRF; adding the local prior makes each tissue class more contiguous by encouraging the voxels to belong to the same class as their neighbours; introducing the multi-atlas prior refines the misclassified structures. By combining all the information, KM-MRF-MAS produces the most accurate classification. Most of the misclassification is left at the intersection between two tissues which is probably caused by the misregistration and the partial volume effect (a single voxel contains more than one tissue). From Table 6.8, we can also observe that in contrast with KM

and KM-MRF, MAS achieves better or comparable results with regard to all the metrics. On the basis of the classification accuracy obtained by MAS, the additional intensity and local prior information do not significantly benefit the classification in terms of *AC*, DSC_{GM} and DSC_{WM}. However, the classification accuracy of CSF is significantly improved by combining all three priors. From Fig. 6.3, it can be seen that on that specific slice the differences between MAS and KM-MRF-MAS are negligible. These results indicate that the multi-atlas prior contributes more than the image intensity and local prior in brain tissue classification. Finally, taking all the comparison results into account, the four classification models are ranked as: KM-MRF significantly outperforms KM; MAS performs better than both KM and KM-MRF; the combined model KM-MRF-MAS achieves the best performance.

For MICCAI2013, we do not include the evaluation results of *AC* because this metric is not used by the challenge. Significant differences exist among the four methods in terms of DSC_{CSF}, DSC_{GM} and DSC_{WM}. The results of *post hoc* pairwise comparisons with regard to each metric are consistent: KM-MRF performs significantly better than KM; MAS performs the worst among the four methods; KM-MRF-MAS produced segmentation results comparable to KM-MRF. The poor performance of MAS can probably be associated with the low number of available atlases, with the selected atlases not similar enough to the test image. Because of this, the incorporation of the multi-atlas prior does not significantly improve the tissue classification performance. Thus we can draw the conclusion that introducing the local prior significantly improve the tissue classification; the efficiency of the multi-atlas prior highly depends on the number and relevance of the selected atlases.

Table 6.8 Comparisons of the performance of different tissue classification models on real data

Database	Method	*AC*	DSC_{CSF}	DSC_{GM}	DSC_{WM}
IBSR18	KM	0.68 ± 0.16	0.20 ± 0.08	0.70 ± 0.15	0.82 ± 0.21
	KM-MRF*	$0.72 \pm 0.18^{\star}$	$0.28 \pm 0.14^{\star}$	$0.75 \pm 0.18^{\star}$	0.80 ± 0.20
	MAS**	$0.88 \pm 0.06^{\star\star}$	$0.70 \pm 0.13^{\star\star}$	$0.91 \pm 0.03^{\star\star}$	0.83 ± 0.14
	KM-MRF-MAS***	$0.88 \pm 0.06^{\star\star}$	$0.71 \pm 0.12^{\star\star\star}$	$0.91 \pm 0.03^{\star\star}$	0.82 ± 0.15
IBSR20	KM	0.65 ± 0.16	0.21 ± 0.07	0.68 ± 0.15	$0.74 \pm 0.19^{\star}$
	KM-MRF*	$0.67 \pm 0.18^{\star}$	$0.34 \pm 0.15^{\star}$	$0.72 \pm 0.17^{\star}$	0.66 ± 0.17
	MAS**	$0.82 \pm 0.04^{\star\star}$	$0.57 \pm 0.09^{\star\star}$	$0.86 \pm 0.03^{\star\star}$	$0.76 \pm 0.06^{\star}$
	KM-MRF-MAS***	$0.82 \pm 0.04^{\star\star}$	$0.57 \pm 0.09^{\star\star\star}$	$0.86 \pm 0.03^{\star\star}$	$0.76 \pm 0.06^{\star}$
MICCAI2013	KM*	-	$0.75 \pm 0.04^{\star}$	$0.76 \pm 0.03^{\star}$	$0.84 \pm 0.03^{\star}$
	KM-MRF**	-	$0.76 \pm 0.04^{\star\star}$	$0.80 \pm 0.02^{\star\star}$	$0.87 \pm 0.02^{\star\star}$
	MAS	-	0.71 ± 0.06	0.70 ± 0.07	0.73 ± 0.08
	KM-MRF-MAS**	-	$0.77 \pm 0.04^{\star\star}$	$0.81 \pm 0.02^{\star\star}$	$0.87 \pm 0.02^{\star\star}$

- Each value represents the mean accuracy and the standard deviation of applying each method on the database.
- The $\star$s at the top right of each value indicate the significant differences of the corresponding result compared with all the others derived by applying other methods on the same database and measured by the same metric. The $\star$s at the top right of each method indicate the significant differences of the performance of the corresponding method compared with all the other methods. The result or method labelled with more $\star$s is significantly better than those with fewer or no $\star$s and the results or methods labelled with equal number of $\star$s have no significant difference between each other.

6.4 Discussion

As described above, the image intensity prior can be utilised by applying simple clustering algorithms like KM for initial classification. However, the performance really depends on the quality of the images. When the noise and intensity inhomogeneity levels are low, the tissues are classified with high accuracy. However, when the images contain high levels of artefacts, the classification can severely deteriorate. In the experiments, we found that KM provides acceptable initial results for most images in the real databases. However, it also produces outliers, producing classification results that are significantly different from the ground truth.

Introducing the local prior significantly benefits the tissue classification in particular when severe noise and intensity inhomogeneity exist in the images. For the real data, the local prior has a positive effect on the classification of CSF and GM and the overall accuracy is significantly improved. The size of the local neighbourhood is data dependent and an appropriate estimation methodology is discussed in Section 6.3.4.

The multi-atlas prior is derived by registering multiple atlas images to the target image and propagating the label maps of the atlases to the target image. The classification is performed by fusing the propagated label maps using a label fusion strategy. The most relevant atlases are selected by measuring the similarities between the atlas images and the target image so that the classification is not biased by irrelevant atlases. In our experiments on the IBSR databases, the atlases were taken from the same database. We tried using atlases from both databases. However, the results showed that the best segmentation is obtained by using the atlases from the same database because of the higher similarity among images. As stated in Section 6.3.5, either the classification performance is significantly improved or the optimal number of atlases to be fused is dramatically decreased by applying the atlas selection. In addition to atlas selection, a sufficient number of available atlases is crucial for the efficiency of the multi-atlas prior. The preprocessing, such as bias field correction and spatial normalisation, applied to the original IBSR databases could contribute to the similarity between images. However, the large number of available atlases greatly increases the chances of selecting the most effective ones for the segmentation of the test images, which significantly benefit MAS. This has been confirmed by the evaluation on the MICCAI2013 database, where the effect of the multi-atlas prior deteriorates severely due to the low number of atlases provided.

The effects of the three image priors on brain tissue classification are compared by comparing the performance of four methods modelling them: KM, KM-MRF, MAS and KM-MRF-MAS. We did not include MRF-MAS because it is not straightforward

to combine the local prior and the multi-atlas prior. In the literature, MRF is often used in combination with a Gaussian mixture model (GMM) [286, 80]. In KM-MRF, KM is used to estimate the initial parameters of the GMM. If we combine MRF with MAS, the only viable solution is replacing KM with MAS, which could result in a more precise estimation for the GMM parameters. However, earlier experiments revealed that a better estimation for the parameters does not improve the segmentation performance, which is confirmed in the literature [286]. In a GMM-MRF based approach, an initial estimation generated from KM is good enough for the segmentation. The effectiveness of the multi-atlas prior is unlikely to be affected by replacing KM with MAS.

From Section 6.3.6, the method using the multi-atlas prior (MAS) achieves better or equal classification compared with those constructed based on the image intensity and the local prior only (KM and KM-MRF) in terms of all metrics. It performs comparably to the combined model (KM-MRF-MAS) with regard to the overall accuracy and the classification accuracies of GM and WM. These results indicate the better effectiveness of the multi-atlas prior in tissue classification compared with the image intensity and the local prior.

We we did not consider the validity of the ground truth provided by the IBSR databases since they have been widely used for the validation of brain image analysis techniques. It has been pointed out that the sulcal CSF voxels in the IBSR datasets are considered as GM, which could affect the tissue classification accuracies measured by DSC [290]. In spite of the uncertainty existing in the ground truth of CSF and GM, we do not think it affects the conclusions we have drawn. First, the improvement of the classification by introducing the local prior has also been validated on two other databases: BrainWeb (see Fig. 6.1) and MICCAI2013 (see Table 6.8), for which we assume the ground truth provided is accurate. Second, the deviation in the ground truth of the IBSR datasets should not affect the performance of the multi-atlas prior. As described in Section 6.2.3, in a multi-atlas based segmentation method, the annotations of the images are actually used to achieve the segmentation. If the annotation is corrected, the segmentation will be adjusted accordingly.

The usage of the multi-atlas prior results in superior performance, but multiple atlases are not always available since human annotation takes great effort. The large amount of registration required in the atlas selection also restricts its application. A compromise was presented [6] and employed [25], which performs the atlas selection after the affine registration and applies the nonrigid registration to the top ranked atlases only. Moreover, registration between images might fail or the test images could not be similar enough to the images of atlases, in which case outliers will be included. A keypoint transfer segmentation approach was proposed to segment

abdominal organs in computerised tomography (CT) images, which propagates the label maps according to the transformation calculated from the matched keypoints in the atlas and test images [303]. This approach requires no registration and yields a segmentation accuracy which compares favourably to that of state-of-the-art methods.

In addition to using the annotated images as atlases and performing the tissue classification in a MAS approach, we can also extract more advanced features such as Gaussian scale-space features [293, 294], Gaussian derivative features [294] or 3D Haar-like features [307] from all the images and the classification is then achieved by training a classifier, such as a k-nearest neighbour classifier [293], a support vector machine [294] or a random forest [307], using these features. Some work has been proposed to extract the features from the tissue probability maps derived from classification results [307, 199] and these features together with those extracted from the images are used for training the subsequent classifiers in a multi-stage tissue classification framework. Hence the tissue probability maps are refined at each classification stage. These works provide state-of-the-art accuracy on the MICCAI2013 database [195]. Unfortunately, the multi-atlas based methods (MAS and KM-MRF-MAS) discussed in this paper do not perform efficiently on this database because of the low number of atlases provided. Images taken from other modalities such as diffusion or T2-weighted MR imaging for the same subject provide more information for brain tissue classification [302]. Longitudinal data can also facilitate tissue classification by taking temporal smoothness into account.

Furthermore, the classification model resulting from applying the classifiers as described above can also be combined with the local and the multi-atlas priors, which are investigated in our paper, to further improve the classification. It has been stated that the classification derived from applying the multi-stage random forest can be combined with the anatomically-constrained multi-atlas segmentation approach [308] to reduce the possible anatomical errors [307]. Conversely, the multi-atlas based approach can be applied first and the trained classifiers are used to refine the classification at 'ambiguous' voxels [125]. Alternatively, an appearance model can be obtained from training classifiers and then combined with the spatial model and the interaction potential, which are generated from the multi-atlas segmentation and the neighbouring information, respectively [293], for the final tissue classification.

6.5 Conclusions

This paper has investigated the effects of three main image priors including intensity, local and multi-atlas priors on the tissue classification in 3D T1-weighted brain MR images. The modelling of these image priors, the combination of the models constructed using one or two image priors and the tissue classification approach of each individual or combined model have been described. The performance of all the models using the three image priors has been validated on the simulated and real data. We have also discussed the impacts of varying the key parameters or factors of each model thoroughly. The classification results have been evaluated based on several metrics and the performance of all the models has been compared quantitatively and qualitatively.

The image intensity utilised by KM generates acceptable initial classification for most images to form the basis for further processing. However, its performance deteriorates from severe artefacts and it produces a few outliers in real data. Introducing the local prior overcomes the disadvantage of KM to some extent and provides an incremental performance improvement. The multi-atlas prior is beneficial in refining the misclassified structures. It contributes more in brain tissue classification than the image intensity and the local prior. Appropriate atlas selection and sufficient number of available atlases are crucial in MAS based approaches. The local weighting method performs better than the global methods in multi-atlas label fusion. The model combining all three image priors achieves better or equal results compared to the other approaches.

Chapter 7

Conclusions

In this chapter, we summarise the content of the thesis, list the main contributions of our work and propose future directions.

7.1 Summary of the Thesis

Automatic and robust 3D brain image analysis can assist in disease diagnosis, surgical planning and patient follow-up. Among all the imaging techniques, MRI has drawn great attention due to its non-ionization, high spatial resolution and excellent soft tissue contrast. Brain MR image segmentation (e.g. the whole brain, structures, tissue or lesion segmentation) could benefit voxel based morphometry, surface reconstruction or 3D visualisation. Brain image registration is used for multi-modality image fusion, longitudinal analysis and population modelling. In this thesis, we have introduced the clinical background of 3D brain MR image analysis and reviewed existing approaches for image segmentation, registration and evaluation. For the most commonly used methods for brain segmentation and brain image registration, we have identified their disadvantages and proposed possible improvements. For brain tissue classification, we have described the modelling of three commonly used image based features, analysed the key parameters involved in each model and compared the effectiveness of these features in tissue classification.

In the clinical background (Chapter 2), we have briefly described a wide range of brain imaging technologies and compared them in terms of imaging principle, applications, advantages and disadvantages. Then we have provided a detailed introduction of brain MRI including its physical principle, multiple sequences and various image views. A variety of applications of 3D brain MR image analysis with regard to segmentation, registration and classification have been presented. Existing challenges

in this area, including image noise, intensity inhomogeneity and partial volume effect have also been described. The databases used for our experiments, including one simulated and four real databases, have also been introduced.

A large number of methods have been proposed for image segmentation and registration and we have conducted a thorough technical literature review in Chapter 3 on generic image segmentation, registration approaches and their evaluation measures. For image segmentation, we divided existing methods into six categories, from the basic techniques such as thresholding and region growing to the more advanced approaches such as deformable models, machine learning algorithms and atlas based methods. For the methods belonging to each category, we also subdivided them into several subgroups, described their working principles, listed a few applications in medical image analysis and presented their advantages and disadvantages. Image segmentation accuracy can be objectively evaluated by overlap, distance and probabilistic measures. The examples of these measures have been formulated. For image registration, we divided the process into four main procedures: feature mapping, transformation estimation, parameter optimisation and image resampling. The methods commonly used at each stage have also been reviewed. The performance of image registration can be evaluated by volume overlap, surface overlap, boundary distance, volume similarity and transformation consistency measures. According to the literature, it was demonstrated that the metrics based on local regions are more reliable than those defined on the whole volume or transformations. Finally, two statistical analysis techniques, paired-sample t-test and MANOVA, used to compare the performance of different methods have been described. In our work, paired-sample t-test was used to analyse the impact of the parameter on the performance of the method with respect to one metric or compare the performance of two methods measured by the same metric; MANOVA was used for comparing the performance of several methods regarding to several metrics.

Brain segmentation is a prerequisite of subsequent image analysis such as registration or tissue classification. Among all the methods presented in the literature, BET is the most widely used because of its simplicity, accuracy and insensitivity to parameter settings. It works based on the principle of the deformable models, which use internal forces to ensure the smoothness of the evolving brain surface and the external forces to interact with the image. However, BET usually requires a large number of iterations for the deforming brain surface to converge; it lacks the strategy to deal with the surface self-intersection effectively; and it tends to undersegment the brain for some databases. To overcome these drawbacks, we proposed an improved BET method in Chapter 4, which improved the displacement of each vertex on the evolving surface and embedded

an independent surface reconstruction step at each iteration. An additional search path was used for the method to perform more robustly at the bottom of the brain. Moreover, when the surface deforming was completed, a FCM based process was applied to reduce the false positives. All these strategies lead to faster convergence and significantly improved segmentation results. Besides the original BET, the improved method has also been qualitatively and quantitatively compared with two other popular methods, showing competitive segmentation performance.

Brain image registration is an important preprocessing step for brain image analysis. Numerous brain image registration approaches have been proposed in the literature, varying in features used, transformation model, similarity measure, optimisation method and image resampling strategy. In terms of features, image intensity, landmarks either manually annotated or automatically detected, curves, surfaces or volumes can be used for brain image registration. In terms of the transformation model, due to the complex brain internal structures, a nonrigid transformation is needed to match two brain images and rigid/affine transformations are often used for preregistration. In the process of registration, a multiresolution scheme is usually adopted for faster implementation and more accurate registration. Demons is one of the most commonly used registration method, which provides a good tradeoff between the registration accuracy and execution speed. The classic demons method uses the whole image grid as demon points, at which the driving forces are calculated to drive the registration process. Considering the edges, especially salient edges, are more important than the other points in an image, we investigated the effect of using salient edges as demon points in demons registration in Chapter 5. The edges were detected by extending Lindeberg's scale-space theory into 3D space and the saliency of each edge point was calculated. A series of experiments were conducted to compare the registration performance of demons between using the whole image and using all the detected edges, gradually decreased number of salient edges at each resolution or gradually decreased number of salient edges at reduced levels of resolutions. We also compared demons registration performance using various demons selection mechanisms such as non-salient edges, randomly sampled points, randomly sampled edges, etc. It has been demonstrated that salient edges contribute more than either non-salient edges or randomly sampled points in demons registration; statistically equal performance is obtained by demons registration using salient edges compared with that using the whole image as demon points, while the execution time is dramatically reduced.

Brain tissue classification could be beneficial for quantitative analysis, image registration, lesion segmentation and cortical surface extraction. A variety of features have been used for brain tissue classification in the literature. In Chapter 6, we

investigated the effects of three main image information priors including intensity, local prior and multi-atlas prior on the performance of tissue classification. The modelling of each prior, the combination of the models constructed from one or two priors and the tissue classification approach of each individual or combined model have been described in detail. Image intensity could be used by k-means (KM) to obtain preliminary brain tissue classification. The local prior was modelled by a Markov random field (MRF), which could be combined with KM (KM-MRF) by modelling each resulting tissue cluster from KM with a Gaussian distribution. The multi-atlas prior could be used individually in Multi-atlas segmentation (MAS) to perform tissue classification. Results indicated that atlas selection plays an important role in MAS by selecting the most relevant atlases. By comparing the performance of four multi-atlas label fusion methods, it has been shown that the local label fusion strategy performs better than global ones. Finally, the multi-atlas prior was combined with image intensity and the local prior (KM-MRF-MAS) into a Bayesian framework, in which tissue classification was achieved by applying an EM approach. The impacts of the key parameters in KM-MRF, MAS and KM-MRF-MAS on tissue classification accuracy have been analysed. The performance of four models (KM, KM-MRF, MAS and KM-MRF-MAS) has been compared quantitatively and qualitatively on both simulated and real data. From the experimental results, KM using the image intensity generates reasonable tissue classification for most images. However, its performance could be significantly deteriorated by the artefacts contained in the images and it produces a few outliers on real brain image data. Introducing the local prior overcomes this weakness to some extent. The multi-atlas prior is beneficial in refining the misclassified brain structures, which leads to significant improvement of the classification accuracy. In contrast to the image intensity and the local prior, the multi-atlas prior possesses the best effectiveness in brain tissue classification at the cost of high computation demand, which is caused by multi-atlas registration. The method combining all three priors (KM-MRF-MAS) obtains the best classification performance with respect to some of the metrics.

7.2 Contributions

The contributions of our work presented in this thesis are:

- A wide range of brain imaging modalities were compared in terms of physical principles, applications, advantages and disadvantages.

- An extensive literature review on generic image segmentation, registration techniques and their evaluation measures was conducted. Both traditional and state-of-the-art image segmentation and registration technologies were described. From a thorough review, we were able to find out the most appropriate metrics to evaluate our brain MR image segmentation and registration methods.
- An improved BET method for brain segmentation was proposed, which overcomes weaknesses existing in the original method and results in faster algorithm convergence and significantly improved segmentation performance.
- We extended Lindeberg's scale-space theory into 3D to extract salient edges from brain MR images.
- A salient edge guided brain image registration method using demons was developed. Compared with the classic demons, which uses the whole brain image as demon points in the registration process, our method using the salient edges produces statistically equal results while substantially saving in computational time.
- A comparative study of brain tissue classification using image information priors was conducted. The effects of three main image information priors on brain tissue classification were investigated. We mathematically formulated their modelling, analysed the key parameters or factors involved in each model and compared the performance of each model on brain tissue classification. This work provides a general guide on what image based features can be used for effective brain tissue classification.
- The performance of global, local multi-atlas label fusion strategies and their combination was compared for brain tissue classification using multi-atlas segmentation methods.
- In the comparison of the performance of different methods evaluated by several metrics, the repeated measures MANOVA was applied to analyse the differences in experimental results in terms of the combination of all metrics or each individual metric.

7.3 Future Work

Based on the work presented in this thesis, a list of future directions are proposed as follows:

- A comparison between our salient edge guided demons registration method and other popular brain image registration approaches reviewed in Section 5.1.1 (e.g. B-splines, HAMMER, CVS, etc) could be conducted.

- Atlas selection plays an important role in multi-atlas brain tissue classification. We could select the relevant atlases after affine registration rather than after nonrigid registration, which might alleviate the burden caused by multi-atlas registration in this step.

- Furthermore, a keypoint transfer segmentation approach proposed in [303] might be modified and applied for multi-atlas brain tissue segmentation. This approach might completely avoid the registration while maintaining the high accuracy of multi-atlas segmentation.

- In addition to the image information priors investigated in our work for brain tissue classification, other more advanced features such as Gaussian scale-space features [293, 294] or 3D Haar-like features [307] can be detected from multiple atlases. These features can be used by machine learning algorithms for brain tissue classification.

- Deep learning algorithms [14, 86] have shown impressive performance in image segmentation [272, 288]. We can learn more representative features from brain MR images by applying these techniques to brain tissue classification.

- The machine learning algorithms together with various image features can also be applied to brain pathological region segmentation such as tumours or MS lesions segmentation.

- In our work, we only used T1-w brain MR images. In the future, we could incorporate information from multi-modality images for conducting brain image analysis.

Appendix A

Calculation of Multi-atlas Prior Energy

The multi-atlas prior energy $E_{Apr}(x)$ in Eq. 6.21 is formulated as

$$\begin{aligned} E_{Apr}(x) &= -\ln P_{Apr}(x) \\ &= -\ln \prod_{i \in \mathscr{S}} P_{Apr}(x_i) \\ &= -\sum_{i \in \mathscr{S}} \ln P_{Apr}(x_i). \end{aligned} \tag{A.1}$$

For each tissue class l, the multi-atlas prior probability $P_{Apr}(x_i = l)$ based on each label fusion strategy is calculated as:

MV:

$$\frac{\sum_{a \in \mathscr{A}} \xi_{x_i^a, l}}{|\mathscr{A}|}$$

where $|\mathscr{A}|$ is the number of atlases in $\mathscr{A}$;

MI:

$$\frac{\sum_{a \in \mathscr{A}} \frac{I(a;y)}{I_{max}} \xi_{x_i^a, l}}{\sum_{a \in \mathscr{A}} \frac{I(a;y)}{I_{max}}};$$

PPBM:

$$\frac{\sum_{a \in \mathscr{A}} \sum_{j \in \mathscr{M}_i} P(y_i|j,a) P(x_i = l|j,x^a)}{\sum_{u \in \mathscr{L}} \sum_{a \in \mathscr{A}} \sum_{j \in \mathscr{M}_i} P(y_i|j,a) P(x_i = u|j,x^a)};$$

MI-PPBM:

$$\frac{\sum_{a \in \mathscr{A}} \sum_{j \in \mathscr{M}_i} \frac{I(a;y)}{I_{max}} P(y_i|j,a) P(x_i = l|j,x^a)}{\sum_{u \in \mathscr{L}} \sum_{a \in \mathscr{A}} \sum_{j \in \mathscr{M}_i} \frac{I(a;y)}{I_{max}} P(y_i|j,a) P(x_i = u|j,x^a)}.$$

Appendix B

Glossary of Terms

Acronyms

3**D**	Three dimensional
*k***NN**	*k*-nearest neighbours
ADC	apparent diffusion coefficient
ANNs	Artificial neural network
ASSD	Average symmetric surface distance
AUC	area under the ROC curve
BET	Brain Extraction Tool
BSE	Brain Surface Extractor
CNNs	Convolutional Neural Networks
CSF	cerebral spinal fluid
CT	Computed Tomography
CVS	combining volumetric and surface
D-w	Diffusion-weighted
DBNs	Deep Belief Networks
DL	deep learning
DNNs	Deep neural networks
DSC	Dice similarity coefficient
EEG	Electroencephalogram

EM	expectation maximisation
FCM	fuzzy *c*-means
FLAIR	Fluid attenuation inversion recovery
fMRI	Functional Magnetic Resonance Imaging
FN	false negatives
FNR	false negative rate
FP	false positives
FPR	false positive rate
GM	grey matter
GMM	Gaussian mixture models
HAMMER	hierarchical attribute matching mechanism for elastic registration
HD	Hausdorff distance
HWA	Hybrid Watershed Algorithm
IBSR	The Internet Brain Segmentation Repository
IR	inversion recovery
KM	*k*-means
LPBA40	The LONI Probabilistic Atlas
MANOVA	Multivariate Analysis of Variance
MAP	maximum a posterior
MAS	multi-atlas segmentation
MASD	Mean average surface distance
MD	Mean distance
MEG	Magnetoencephalography
MI	Mutual Information
MICCAI	Medical Image Computing and Computer-Assisted Intervention
MO	Mean overlap
MR	Magnetic Resonance

MRF Markov random field
MRI Magnetic Resonance Imaging
MS multiple sclerosis
MTP misclassified tissue percentage
MV Majority Voting

NCC normalised cross-correlation
NMI normalised mutual information

PD-w proton density weighted
PEF probabilistic extra fraction
PET Positron Emission Tomography
POF probabilistic overlap fraction
PPBM Probabilistic Patch-Based Model
PSI probabilistic similarity index

RBM Restricted Boltzmann Machine
RF Random forests
RMS root of mean squares
ROC Receiver Operating Characteristic

SEN Sensitivity
SPE Specificity
SPECT Single Photon Emission Computed Tomography
SSDA sequential similarity detection algorithm
SVM Support vector machine

T1-w T1-weighted
T2-w T2-weighted
TE echo time
TN true negatives
TO Target overlap
TP true positives
TR repetition time

UO Union overlap

US	Ultrasound Imaging
WM	white matter

Appendix C

List of Publications

The following papers are the publications related to this thesis during my PhD study:

- **Liping Wang**, Reyer Zwiggelaar. Salient edge guided 3D brain MR image registration using Demons. Proceedings of the 19^{th} Conference on Medical Image Understanding and Analysis (MIUA), 2015, pp. 14-19. (Oral)
- **Liping Wang**, Ziming Zeng and Reyer Zwiggelaar. An improved BET method for brain segmentation. Proceedings of 22^{nd} IEEE International Conference on Pattern Recognition (ICPR), 2014, pp. 3221-3226.
- Ziming Zeng, Chunlei Han, **Liping Wang** and Reyer Zwiggelaar. Unsupervised brain tissue segmentation by using bias correction fuzzy c-means and class-adaptive hidden Markov random field modelling. Frontier and Future Development of Information Technology in Medicine and Education, Volume 269 of the series Lecture Notes in Electrical Engineering, 2014, pp. 579-587.
- Zhili Chen, **Liping Wang**, Erika Denton and Reyer Zwiggelaar. A multiscale blob representation of mammographic parenchymal patterns and mammographic risk assessment. 15^{th} International Conference on Computer Analysis of Images and Patterns, Volume 8048 of the series Lecture Notes in Computer Science, 2013, pp. 346-353.

Paper under review:

- **Liping Wang** and Reyer Zwiggelaar. A comparative study of brain tissue classification using different image based features. Submitted to Medical Physics.

References

[1] Sherif Abdelsayed, Dan Ionescu, and David Goodenough. Matching and registration method for remote sensing images. In *Geoscience and Remote Sensing Symposium, 1995. 'Quantitative Remote Sensing for Science and Applications', International*, volume 2, pages 1029–1031. IEEE, 1995.

[2] Guillermo N Abras and Virginia Laura Ballarín. A weighted k-means algorithm applied to brain tissue classification. *Journal of Computer Science and Technology*, 5, 2005.

[3] Rolf Adams and Leanne Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.

[4] S Agatonovic-Kustrin and R Beresford. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, 22(5):717–727, 2000.

[5] Mohamed N Ahmed, Sameh M Yamany, Nevin Mohamed, Aly A Farag, and Thomas Moriarty. A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 21(3): 193–199, 2002.

[6] Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Joseph V Hajnal, and Daniel Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage*, 46(3):726–738, 2009.

[7] Daniel G Amen and Blake D Carmichael. High-resolution brain spect imaging in adhd. *Annals of Clinical Psychiatry*, 9(2):81–86, 1997.

[8] Nina Amenta, Sunghee Choi, and Ravi Krishna Kolluri. The power crust. In *Proceedings of the Sixth ACM Symposium on Solid Modeling and Applications*, pages 249–266, New York, USA, 2001. ACM.

[9] Petronella Anbeek, Koen L Vincken, Glenda S Van Bochove, Matthias JP Van Osch, and Jeroen van der Grond. Probabilistic segmentation of brain tissue in MR imaging. *Neuroimage*, 27(4):795–804, 2005.

[10] Petronella Anbeek, Koen L Vincken, Floris Groenendaal, Annemieke Koeman, Matthias JP Van Osch, and Jeroen Van der Grond. Probabilistic brain tissue segmentation in neonatal magnetic resonance imaging. *Pediatric research*, 63 (2):158–163, 2008.

[11] Elena Antonova, Tonmoy Sharma, Robin Morris, and Veena Kumari. The relationship between brain structure and neurocognition in schizophrenia: a selective review. *Schizophrenia Research*, 70(2):117–145, 2004.

[12] Nur Arad, Nira Dyn, Daniel Reisfeld, and Yehezkel Yeshurun. Image warping by radial basis functions: application to facial expressions. *CVGIP: Graphical Models and Image Processing*, 56(2):161–172, 1994.

[13] Babak A Ardekani, Stephen Guckemus, Alvin Bachman, Matthew J Hoptman, Michelle Wojtaszek, and Jay Nierenberg. Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans. *Journal of Neuroscience Methods*, 142(1):67–76, 2005.

[14] Itamar Arel, Derek C Rose, and Thomas P Karnowski. Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *Computational Intelligence Magazine, IEEE*, 5(4):13–18, 2010.

[15] Peter Armitage, Geoffrey Berry, and John NS Matthews. *Statistical methods in medical research*. John Wiley & Sons, 2008.

[16] Xabier Artaechevarria, Arrate Munoz-Barrutia, and Carlos Ortiz-de Solórzano. Combination strategies in multi-atlas image segmentation: Application to brain mr data. *IEEE Transactions on Medical Imaging*, 28(8):1266–1277, 2009.

[17] John Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113, 2007.

[18] John Ashburner and Karl J Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4):254–266, 1999.

[19] John Ashburner and Karl J Friston. Unified segmentation. *NeuroImage*, 26(3): 839–851, 2005.

[20] John Ashburner, John G Csernansk, Christos Davatzikos, Nick C Fox, Giovanni B Frisoni, and Paul M Thompson. Computer-assisted imaging to assess brain structure in healthy and diseased brains. *The Lancet Neurology*, 2(2): 79–88, 2003.

[21] Andrew J Asman and Bennett A Landman. Non-local statistical label fusion for multi-atlas segmentation. *Medical Image Analysis*, 17(2):194–208, 2013.

[22] Alzheimer's Association. 2011 Alzheimer's disease facts and figures. *Alzheimer's & Dementia: the Journal of the Alzheimer's Association*, 7(2): 208, 2011.

[23] Scott W Atlas. *Magnetic resonance imaging of the brain and spine*, volume 1. Lippincott Williams & Wilkins, 2009.

[24] Guillaume Auzias, Olivier Colliot, Joan Alexis Glaunes, Matthieu Perrot, Jean-François Mangin, Alain Trouve, and Sylvain Baillet. Diffeomorphic brain registration under exhaustive sulcal constraints. *IEEE Transactions on Medical Imaging*, 30(6):1214–1227, 2011.

[25] Wenjia Bai, Wenzhe Shi, Declan P O'Regan, Tong Tong, Haiyan Wang, Shahnaz Jamil-Copley, Nicholas S Peters, and Daniel Rueckert. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. *IEEE Transactions on Medical Imaging*, 32(7):1302–1315, 2013.

[26] Caroline Baillard, Pierre Hellier, and Christian Barillot. Segmentation of brain 3D MR images using level sets and dense registration. *Medical image analysis*, 5(3):185–194, 2001.

[27] Rohit Bakshi, Suzie Ariyaratana, Ralph HB Benedict, and Lawrence Jacobs. Fluid-attenuated inversion recovery magnetic resonance imaging detects cortical and juxtacortical multiple sclerosis lesions. *Archives of Neurology*, 58(5):742–748, 2001.

[28] Serdar K Balci, Polina Golland, Martha Shenton, and William M Wells. Free-form B-spline deformation model for groupwise registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 10, page 23. NIH Public Access, 2007.

[29] Daniel I Barnea and Harvey F Silverman. A class of algorithms for fast digital image registration. *IEEE Transactions on Computers*, 100(2):179–186, 1972.

[30] Harry G Barrow, Jay M Tenenbaum, Robert C Bolles, and Helen C Wolf. Parametric correspondence and chamfer matching: Two new techniques for image matching. Technical report, DTIC Document, 1977.

[31] Stefan Bauer, Lutz-P Nolte, and Mauricio Reyes. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pages 354–361. Springer, 2011.

[32] Pierre-Louis Bazin and Dzung L Pham. Statistical and topological atlas based brain image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pages 94–101. Springer, 2007.

[33] Pierre-Louis Bazin and Dzung L Pham. Topology-preserving tissue classification of magnetic resonance brain images. *IEEE transactions on medical imaging*, 26(4):487–496, 2007.

[34] Pierre-Louis Bazin and Dzung L Pham. Homeomorphic brain image segmentation with topological and statistical atlases. *Medical Image Analysis*, 12(5): 616–625, 2008.

[35] Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302, 1986.

[36] Paul J Besl and Neil D McKay. Method for registration of 3-D shapes. In *Robotics-DL Tentative*, volume 14, pages 586–606. International Society for Optics and Photonics, 1992.

[37] James C Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[38] Kanwal K Bhatia, Joseph V Hajnal, Basant K Puri, A David Edwards, and Daniel Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004*, pages 908–911. IEEE, 2004.

[39] Richard Bitar, General Leung, Richard Perng, Sameh Tadros, Alan R Moody, Josee Sarrazin, Caitlin McGregor, Monique Christakis, Sean Symons, Andrew Nelson, and Timothy P Roberts. MR pulse sequences: What every radiologist wants to know but is afraid to ask 1. *Radiographics*, 26(2):513–537, 2006.

[40] Fred L Bookstein. Thin-plate splines and the atlas problem for biomedical images. In *Information Processing in Medical Imaging*, pages 326–342. Springer, 1991.

[41] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152. ACM, 1992.

[42] Yuri Boykov and Gareth Funka-Lea. Graph cuts and efficient ND image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.

[43] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[44] Ron Bracewell. The Fourier transform and its applications. *New York*, 5, 1965.

[45] Andrew P Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

[46] BrainWeb. BrainWeb: Simulated Brain Database. http://brainweb.bic.mni.mcgill.ca/brainweb/. [accessed May 8, 2013].

[47] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[48] Morten Bro-Nielsen and Claus Gramkow. Fast fluid registration of medical images. In *Visualization in Biomedical Computing*, pages 265–276. Springer, 1996.

[49] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Computing Surveys (CSUR)*, 24(4):325–376, 1992.

[50] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. Image denoising methods. a new nonlocal principle. *Society for Industrial and Applied Mathematics (SIAM) Review*, 52(1):113–147, 2010.

[51] Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Meritxell Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 104(3):e158–e177, 2011.

[52] Pascal Cachier, Xavier Pennec, and Nicholas Ayache. Fast non rigid matching by gradient descent: study and improvements of the'demons' algorithm. *Research Report-3706*, 1999.

[53] Nathan D Cahill, J Alison Noble, and David J Hawkes. A demons algorithm for image registration with locally adaptive regularization. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pages 574–581. Springer, 2009.

[54] Weiling Cai, Songcan Chen, and Daoqiang Zhang. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Pattern Recognition*, 40(3):825–838, 2007.

[55] Robert L Cannon, Jitendra V Dave, and James C Bezdek. Efficient implementation of the fuzzy c-means clustering algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2):248–255, 1986.

[56] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):679–698, 1986.

[57] Aaron Carass, Jennifer Cuzzocreo, M Bryan Wheeler, Pierre-Louis Bazin, Susan M Resnick, and Jerry L Prince. Simple paradigm for extra-cerebral tissue removal: algorithm and analysis. *NeuroImage*, 56(4):1982–1992, 2011.

[58] Jaime S Cardoso and Luís Corte-Real. Toward a generic evaluation of image segmentation. *IEEE Transactions on Image Processing*, 14(11):1773–1782, 2005.

[59] Neil R Carlson. *Physiology of Behavior 11th Edition*. Pearson, 2012.

[60] Sandeep Chaplot, LM Patnaik, and NR Jagannathan. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomedical Signal Processing and Control*, 1(1):86–92, 2006.

[61] Tao Chen, Kai-Kuang Ma, and Li-Hui Chen. Tri-state median filter for image denoising. *IEEE Transactions on Image Processing*, 8(12):1834–1838, 1999.

[62] Ming-Chang Chiang, Rebecca A Dutton, Kiralee M Hayashi, Oscar L Lopez, Howard J Aizenstein, Arthur W Toga, James T Becker, and Paul M Thompson. 3D pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry. *NeuroImage*, 34(1):44–60, 2007.

[63] Gary E Christensen. Consistent linear-elastic transformations for image matching. In *Information Processing in Medical Imaging*, pages 224–237. Springer, 1999.

[64] Gary E Christensen, Xiujuan Geng, Jon G Kuhl, Joel Bruss, Thomas J Grabowski, Imran A Pirwani, Michael W Vannier, John S Allen, and Hanna Damasio. Introduction to the non-rigid image registration evaluation project (NIREP). *Lecture Notes in Computer Science*, 4057:128–135, 2006.

[65] Dan Ciresan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in Neural Information Processing Systems*, pages 2843–2851, 2012.

[66] Chris A Cocosco, Alex P Zijdenbos, and Alan C Evans. A fully automatic and robust brain MRI tissue classification method. *Medical image analysis*, 7(4): 513–527, 2003.

[67] Laurent D Cohen. On active contour models and balloons. *CVGIP: Image Understanding*, 53(2):211–218, 1991.

[68] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th International Conference on Machine Learning*, pages 160–167. ACM, 2008.

[69] Gregory F Cooper and Edward Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4):309–347, 1992.

[70] Timothy F Cootes, Andrew Hill, Christopher J Taylor, and Jane Haslam. The use of active shape models for locating structures in medical images. In *Information Processing in Medical Imaging*, pages 33–47. Springer, 1993.

[71] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[72] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):681–685, 2001.

[73] Nicolas Cordier, Bjoern Menze, Hervé Delingette, and Nicholas Ayache. Patch-based segmentation of brain tissues. In *MICCAI Challenge on Multimodal Brain Tumor Segmentation*, pages 6–17. IEEE, 2013.

[74] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[75] Ingemar J Cox, Satish B Rao, and Yu Zhong. Ratio regions: a technique for image segmentation. In *Proceedings of the 13th International Conference on Pattern Recognition, 1996*, volume 2, pages 557–564. IEEE, 1996.

[76] Cranial Untrasound. Cranial Ultrasound/Head Ultrasound. http://www.radiologyinfo.org/en/info.cfm?pg=ultrasound-cranial. [accessed April 5, 2016].

[77] William R Crum, Thomas Hartkens, and DLG Hill. Non-rigid image registration: theory and practice. *The British Journal of Radiology*, 2014.

[78] CT. Radiology masterclass: the home of radiology education for medical students and health care professionals. http://www.radiologymasterclass.co.uk/gallery/ct_brain/ct_brain_stacks/occipital_infarct_ct_brain.html. [accessed January 26, 2016].

[79] Meritxell Bach Cuadra, Claudio Pollo, Anton Bardera, Olivier Cuisenaire, J-G Villemure, and J Thiran. Atlas-based segmentation of pathological MR brain images using a model of lesion growth. *IEEE Transactions on Medical Imaging*, 23(10):1301–1314, 2004.

[80] Meritxell Bach Cuadra, Leila Cammoun, Torsten Butz, Olivier Cuisenaire, and Jean-Philippe Thiran. Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. *IEEE Transactions on Medical Imaging*, 24(12):1548–1565, 2005.

[81] Anders M Dale, Bruce Fischl, and Martin I Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194, 1999.

[82] Christos Davatzikos, Jerry L Prince, and R Nick Bryan. Image registration based on boundary mapping. *IEEE Transactions on Medical Imaging*, 15(1):112–115, 1996.

[83] Renske De Boer, Henri A Vrooman, Fedde Van Der Lijn, Meike W Vernooij, M Arfan Ikram, Aad Van Der Lugt, Monique MB Breteler, and Wiro J Niessen. White matter lesion extension to automatic brain tissue segmentation on MRI. *Neuroimage*, 45(4):1151–1161, 2009.

[84] E De Castro and C Morandi. Registration of translated and rotated images using finite fourier transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (5):700–703, 1987.

[85] Hervé Delingette and Johan Montagnat. Shape and topology constraints on parametric active contours. *Computer Vision and Image Understanding*, 83(2):140–171, 2001.

[86] Li Deng and Dong Yu. Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7(3–4):197–387, 2014.

[87] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[88] Diffusion-Weighted Image. MRIMASTER.com: Diffusion-Weighted Imaging (DWI)/EPI DWI. https://mrimaster.com/characterise%20image%20dwi%20.html. [accessed Febuary 4, 2016].

[89] Helmut Dittmann, Bernhard Matthias Dohmen, Frank Paulsen, Kai Eichhorn, Susanne Martina Eschmann, Marius Horger, Manfred Wehrmann, Hans Juergen Machulla, and Roland Bares. [18F]FLT PET for diagnosis and staging of thoracic tumours. *European Journal of Nuclear Medicine and Molecular Imaging*, 30(10):1407–1412, 2003.

[90] Neil A Dodgson. Quadratic interpolation for image resampling. *IEEE Transactions on Image Processing*, 6(9):1322–1326, 1997.

[91] Michael Donaghy. *Brain's diseases of the nervous system*. Oxford University Press New York, 2009.

[92] Gianluca Donato and Serge Belongie. Approximate thin plate spline mappings. In *Computer Vision—ECCV 2002*, pages 21–31. Springer, 2002.

[93] Jimit Doshi, Guray Erus, Yangming Ou, Bilwaj Gaonkar, and Christos Davatzikos. Multi-atlas skull-stripping. *Academic radiology*, 20(12):1566–1576, 2013.

[94] Jia Du, Laurent Younes, and Anqi Qiu. Whole brain diffeomorphic metric mapping via integration of sulcal and gyral curves, cortical surfaces, and images. *NeuroImage*, 56(1):162–173, 2011.

[95] Ciobanu Dumitru and Vasilescu Maria. Advantages and disadvantages of using neural networks for predictions. *Ovidius University Annals, Series Economic Sciences*, 13(1), 2013.

[96] Klaus-Dietrich Ebel and Gabriele Benz-Bohm. *Differential diagnosis in pediatric radiology*. Thieme, 1999.

[97] EEG. positiveneuro.com: Will the real BCI please stand up? https://loverev.wordpress.com/2010/03/03/will-the-real-bci-please-stand-up/. [accessed January 26, 2016].

[98] RG Emerson and TA Pedley. Clinical neurophysiology: electroencephalography and evoked potentials. *Neurology in Clinical Practice. 6th ed. Philadelphia, PA: Butterworth-Heinemann*, 2012.

[99] Kjell Erlandsson, Irene Buvat, P Hendrik Pretorius, Benjamin A Thomas, and Brian F Hutton. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. *Physics in Medicine and Biology*, 57(21):R119, 2012.

[100] S Esakkirajan, T Veerakumar, Adabala N Subramanyam, and Prem CH Chand. Removal of high density salt and pepper noise through modified decision based unsymmetric trimmed median filter. *IEEE Signal Processing Letters*, 18(5): 287–290, 2011.

[101] Simon F Eskildsen, Pierrick Coupé, Vladimir Fonov, José V Manjón, Kelvin K Leung, Nicolas Guizard, Shafik N Wassef, Lasse Riis Østergaard, and D Louis Collins. Beast: Brain extraction based on nonlocal segmentation technique. *NeuroImage*, 59(3):2362–2373, 2012.

[102] Simon F Eskildsen, Pierrick Coupé, Vladimir Fonov, José V Manjón, Kelvin K Leung, Nicolas Guizard, Shafik N Wassef, Lasse Riis Østergaard, D Louis Collins, Alzheimer's Disease Neuroimaging Initiative, et al. BEaST: brain extraction based on nonlocal segmentation technique. *NeuroImage*, 59(3): 2362–2373, 2012.

[103] Alan C Evans, D Louis Collins, SR Mills, ED Brown, RL Kelly, and Terry M Peters. 3D statistical neuroanatomical models from 305 MRI volumes. In *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.*, pages 1813–1817. IEEE, 1993.

[104] Alexandre X Falcao, Jayaram K Udupa, Supun Samarasekera, and Bruce E Hirsch. User-steered image boundary segmentation. In *Medical Imaging 1996*, pages 278–288. International Society for Optics and Photonics, 1996.

[105] Alexandre X Falcão, Jayaram K Udupa, and Flávio K Miyazawa. An ultra-fast user-steered image segmentation paradigm: live wire on the fly. *IEEE Transactions on Medical Imaging*, 19(1):55–62, 2000.

[106] Qianqian Fang and David A Boas. Tetrahedral mesh generation from volumetric binary and grayscale images. In *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1142–1145, 2009.

[107] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[108] Aaron G Filler. The history, development and impact of computed imaging in neurological diagnosis and neurosurgery: CT, MRI, and DTI. *Nature Proceedings*, 7(1):1–69, 2009.

[109] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, Albert Montillo, Nikos Makris, Bruce Rosen, and Anders

M Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

[110] Jörg Fliege and Benar Fux Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51(3):479–494, 2000.

[111] fMRI. Neuroimaging: Functional magnetic resonance imaging. https://en.wikipedia.org/wiki/Neuroimaging. [accessed January 26, 2016].

[112] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2-3):131–163, 1997.

[113] Guido Gerig, Olaf Kübler, Ron Kikinis, and Ferenc A Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11 (2):221–232, 1992.

[114] David Thomas Gering, W Eric L Grimson, and Ron Kikinis. *Recognizing deviations from normalcy for brain tumor segmentation*. Springer, 2002.

[115] Adam T Gerstenblith and Michael P Rabinowitz. *The Wills eye manual: office and emergency room diagnosis and treatment of eye disease*. Lippincott Williams & Wilkins, 2012.

[116] Ali Gholipour, Nasser Kehtarnavaz, Richard Briggs, Michael Devous, and Kaundinya Gopinath. Brain functional localization: a survey of image registration techniques. *IEEE Transactions on Medical Imaging*, 26(4):427–451, 2007.

[117] Ardeshir Goshtasby and George C Stockman. Point pattern matching using convex hull edges. *IEEE Transactions on Systems, Man and Cybernetics*, (5): 631–637, 1985.

[118] Todd A Gould. How MRI works. *HowStuffWorks, Inc*, 10, 2008.

[119] Leo Grady. Minimal surfaces extend shortest path segmentation methods to 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(2): 321–334, 2010.

[120] Hayit Greenspan, Amit Ruf, and Jacob Goldberger. Constrained gaussian mixture model framework for automatic segmentation of MR brain images. *IEEE transactions on medical imaging*, 25(9):1233–1245, 2006.

[121] Douglas N Greve and Bruce Fischl. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1):63–72, 2009.

[122] Xiao Gu, Ji-Xiang Du, and Xiao-Feng Wang. Leaf recognition based on the combination of wavelet transform and gaussian interpolation. In *Advances In Intelligent Computing*, pages 253–262. Springer, 2005.

[123] Eldad Haber and Jan Modersitzki. Intensity gradient based registration and fusion of multi-modal images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 726–733. Springer, 2006.

[124] William W Hager and Hongchao Zhang. A survey of nonlinear conjugate gradient methods. *Pacific Journal of Optimization*, 2(1):35–58, 2006.

[125] Xiao Han. Learning-boosted label fusion for multi-atlas auto-segmentation. In *International Workshop on Machine Learning in Medical Imaging*, pages 17–24. Springer, 2013.

[126] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, (6):610–621, 1973.

[127] Lei He, Zhigang Peng, Bryan Everding, Xun Wang, Chia Y Han, Kenneth L Weiss, and William G Wee. A comparative study of deformable contour methods on medical image segmentation. *Image and Vision Computing*, 26(2):141–163, 2008.

[128] Rolf A Heckemann, Shiva Keihaninejad, Paul Aljabar, Daniel Rueckert, Joseph V Hajnal, Alexander Hammers, and Alzheimer's Disease Neuroimaging Initiative. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. *NeuroImage*, 51(1):221–227, 2010.

[129] Mattias P Heinrich, Mark Jenkinson, J Michael Brady, and Julia A Schnabel. Non-rigid image registration through efficient discrete optimization. In *Proceeding of Medical Image Understanding and Analysis*, pages 187–192, 2011.

[130] Derek LG Hill, Philipp G Batchelor, Mark Holden, and David J Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46(3):R1, 2001.

[131] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, and Brian Kingsbury. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE*, 29(6):82–97, 2012.

[132] Geoffrey E Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002.

[133] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[134] Daniel P Huttenlocher, Gregory A Klanderman, and William J Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[135] IBSR. NITRC: IBSR. http://www.nitrc.org/projects/ibsr/. [accessed July 1, 2015].

[136] Juan Eugenio Iglesias and Mert R Sabuncu. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*, 24(1):205–219, 2015.

[137] Juan Eugenio Iglesias, Cheng-Yi Liu, Paul M Thompson, and Zhuowen Tu. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Transactions on Medical Imaging*, 30(9):1617–1634, 2011.

[138] Barbara Jesacher, Alain Reimann, Raphael Meier, Roland Wiest, Mauricio Reyes, and Stefan Bauer. Brain tissue segmentation in MRI by supervised classification and regularization. *MICCAI Grand Challenge on MR Brain Image Segmentation workshop*, 2013.

[139] Amod Jog, Snehashis Roy, Jerry L Prince, and Aaron Carass. Mr brain segmentation using decision trees. *Proceedings of the MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS'13)*, 2013.

[140] G Johannsen and J Bille. A threshold selection method using information measures. In *International Conference on Pattern Recognition*, volume 82, pages 140–143, 1982.

[141] Hans J Johnson and Gary E Christensen. Landmark and intensity-based, consistent thin-plate spline image registration. In *Information Processing in Medical Imaging*, pages 329–343. Springer, 2001.

[142] Hans J Johnson and Gary E Christensen. Consistent landmark and intensity-based image registration. *IEEE Transactions on Medical Imaging*, 21(5):450–461, 2002.

[143] B Johnston, M Stella Atkins, B Mackiewich, and M Anderson. Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI. *IEEE Transactions on Medical Imaging*, 15(2):154–169, 1996.

[144] Jagat Narain Kapur, Prasanna K Sahoo, and Andrew KC Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29(3):273–285, 1985.

[145] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[146] David N Kennedy, Pauline A Filipek, and Verne S Caviness Jr. Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 8(1):1–7, 1989.

[147] Ab Rouf Khan, Lei Wang, and Mirza Faisal Beg. Multistructure large deformation diffeomorphic brain registration. *IEEE Transactions on Biomedical Engineering*, 60(2):544–553, 2013.

[148] Satyanad Kichenassamy, Arun Kumar, Peter Olver, Allen Tannenbaum, and Anthony Yezzi Jr. Conformal curvature flows: from phase transitions to active vision. *Archive for Rational Mechanics and Analysis*, 134(3):275–301, 1996.

[149] June Sic Kim, Vivek Singh, Jun Ki Lee, Jason Lerch, Yasser Ad-Dab'bagh, David MacDonald, Jong Min Lee, Sun I Kim, and Alan C Evans. Automated 3D extraction and evaluation of the inner and outer cortical surfaces using a laplacian map and partial volume effect classification. *NeuroImage*, 27(1): 210–221, 2005.

[150] Robert L Kirby and Azriel Rosenfeld. A note on the use of (gray level, local average gray level) space as an aid in threshold selection. Technical report, DTIC Document, 1979.

[151] Arno Klein, Jesper Andersson, Babak A Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E Christensen, D Louis Collins, James Gee, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P Woods, J John Mann, and Ramin V Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802, 2009.

[152] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.

[153] Ralf Kohler. A segmentation system based on thresholding. *Computer Graphics and Image Processing*, 15(4):319–338, 1981.

[154] Efstathios D Kondylis, Thomas A Wozny, Witold J Lipski, Alexandra Popescu, Vincent J DeStefino, Behnaz Esmaeili, Vineet K Raghu, Anto Bagic, and R Mark Richardson. Detection of high-frequency oscillations by hybrid depth electrodes in standard clinical intracranial eeg recordings. *Frontiers in Neurology*, 5, 2014.

[155] N Kovacevic, NJ Lobaugh, MJ Bronskill, B Levine, A Feinstein, and SE Black. A robust method for extraction and automatic segmentation of brain images. *Neuroimage*, 17(3):1087–1100, 2002.

[156] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[157] D-J Kroon and Cornelis H Slump. MRI modalitiy transformation in demon registration. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009*, pages 963–966. IEEE, 2009.

[158] CD Kuglin. The phase correlation image alignment method. In *Proceeding. International Conference on Cybernetics and Society, 1975*, pages 163–165, 1975.

[159] Sai Ho Kwok and Anthony G Constantinides. A fast recursive shortest spanning tree for image segmentation and edge detection. *IEEE Transactions on Image Processing*, 6(2):328–332, 1997.

[160] Jan Kybic, Philippe Thévenaz, and Michael Unser. Multiresolution spline warping for EPI registration. *Proceedings of the SPIE: Mathematical Imaging-Wavelet Applications in Signal and Image Processing, Denver, Colorado*, pages 571–579, 1999.

[161] Wai Lam and Fahiem Bacchus. Learning Bayesian belief networks: An approach based on the MDL principle. *Computational Intelligence*, 10(3):269–293, 1994.

[162] Thomas Robin Langerak, Uulke A van der Heide, Alexis NTJ Kotte, Max A Viergever, Marco van Vulpen, and Josien PW Pluim. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (simple). *IEEE Transactions on Medical Imaging*, 29(12): 2000–2008, 2010.

[163] Thomas M Lehmann, Claudia Gönner, and Klaus Spitzer. Addendum: B-spline interpolation in medical image processing. *IEEE Transactions on Medical Imaging*, 20(7):660–665, 2001.

[164] Louis Lemieux, Georg Hagemann, Karsten Krakow, and Friedrich G Woermann. Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. *Magnetic Resonance in Medicine*, 42(1):127–135, 1999.

[165] Michael E Leventon, W Eric L Grimson, and Olivier Faugeras. Statistical shape influence in geodesic active contours. In *IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings.*, volume 1, pages 316–323. IEEE, 2000.

[166] David N Levin, Charles A Pelizzari, GT Chen, CT Chen, and MD Cooper. Retrospective geometric correlation of MR, CT, and PET images. *Radiology*, 169(3):817–823, 1988.

[167] Jing Li, Xuan Yang, and Jianping Yu. Compact support thin plate spline algorithm. *Journal of Electronics (China)*, 24(4):515–522, 2007.

[168] Alan W-C Liew and Hong Yan. Current methods in the automatic tissue segmentation of 3D magnetic resonance brain images. *Current Medical Imaging Reviews*, 2(1):91–103, 2006.

[169] Zheng Lin, Jesse Jin, and Hugues Talbot. Unseeded region growing for 3D image segmentation. In *Selected papers from the Pan-Sydney workshop on Visualisation-Volume 2*, pages 31–37. Australian Computer Society, Inc., 2000.

[170] Tony Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–156, 1998.

[171] Jia-Xiu Liu, Yong-Sheng Chen, and Li-Fen Chen. Accurate and robust extraction of brain regions using a deformable model based on radial basis functions. *Journal of Neuroscience Methods*, 183(2):255–266, 2009.

[172] Xavier Lladó, Arnau Oliver, Mariano Cabezas, Jordi Freixenet, Joan C Vilanova, Ana Quiles, Laia Valls, Lluís Ramió-Torrentà, and Àlex Rovira. Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches. *Information Sciences*, 186(1):164–185, 2012.

[173] Jyrki MP Lötjönen, Robin Wolz, Juha R Koikkalainen, Lennart Thurfjell, Gunhild Waldemar, Hilkka Soininen, Daniel Rueckert, and Alzheimer's Disease Neuroimaging Initiative. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage*, 49(3):2352–2365, 2010.

[174] Lisa H Lowe and Zachary Bailey. State-of-the-art cranial sonography: Part 1, modern techniques and image interpretation. *American Journal of Roentgenology*, 196(5):1028–1033, 2011.

[175] Florian Luisier, Thierry Blu, and Michael Unser. A new sure approach to image denoising: Interscale orthonormal wavelet thresholding. *IEEE Transactions on Image Processing*, 16(3):593–606, 2007.

[176] Zhen Ma, João Manuel RS Tavares, Renato Natal Jorge, and T Mascarenhas. A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. *Computer Methods in Biomechanics and Biomedical Engineering*, 13(2):235–246, 2010.

[177] James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. Oakland, CA, USA., 1967.

[178] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.

[179] Dwarikanath Mahapatra. Skull stripping of neonatal brain MRI: using prior shape information with graph cuts. *Journal of digital imaging*, 25(6):802–814, 2012.

[180] JB Antoine Maintz and Max A Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[181] Antonios Makropoulos, Christian Ledig, Paul Aljabar, Ahmed Serag, Joseph V Hajnal, A David Edwards, Serena J Counsell, and Daniel Rueckert. Automatic tissue and structural segmentation of neonatal brain MRI using Expectation-Maximization. *MICCAI Grand Challenge: Neonatal Brain Segmentation*, 9: 2012, 2012.

[182] Ravikanth Malladi, James A Sethian, and Baba C Vemuri. Topology-independent shape modeling scheme. In *SPIE's 1993 International Symposium on Optics, Imaging, and Instrumentation*, pages 246–258. International Society for Optics and Photonics, 1993.

[183] José V Manjón, José Carbonell-Caballero, Juan J Lull, Gracián García-Martí, Luís Martí-Bonmatí, and Montserrat Robles. MRI denoising using non-local means. *Medical Image Analysis*, 12(4):514–523, 2008.

[184] Kanti V. Mardia and TJ Hainsworth. A spatial thresholding method for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(6):919–927, 1988.

[185] Stephen Marsland, Carole J Twining, and Chris J Taylor. Groupwise non-rigid registration using polyharmonic clamped-plate splines. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*, pages 771–779. Springer, 2003.

[186] Alberto Martelli. An application of heuristic search methods to edge and contour detection. *Communications of the ACM*, 19(2):73–83, 1976.

[187] Styner Martin, Warfield Simon, Niessen Wiro, van Walsum Theo, Metz Coert, Schaap Michiel, Deng Xiang, Heimann Tobias, and van Ginneken Bram. MS lesion segmentation challenge 2008. http://www.ia.unc.edu/MSseg/index.html. [accessed March 30, 2013].

[188] PM Matthews and P Jezzard. Functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(1):6–12, 2004.

[189] Andrew Mayers. *Introduction to statistics and SPSS in psychology*. Pearson, 2013.

[190] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943.

[191] Tim McInerney and Demetri Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108, 1996.

[192] Tim McInerney and Demetri Terzopoulos. T-snakes: Topology adaptive snakes. *Medical Image Analysis*, 4(2):73–91, 2000.

[193] Andrea Mechelli, Cathy J Price, Karl J Friston, and John Ashburner. Voxel-based morphometry of the human brain: methods and applications. *Current Medical Imaging Reviews*, 1(2):105–113, 2005.

[194] MEG. Mcgovern institute for brain research at MIT. http://mcgovern.mit.edu/news/newsletter/meg-matters/attachment/isik_fig8-copylr/. [Online; accessed January 26, 2016].

[195] Adriënne M Mendrik, Koen L Vincken, Hugo J Kuijf, Marcel Breeuwer, Willem H Bouvy, Jeroen de Bresser, Amir Alansary, Marleen de Bruijne, Aaron Carass, Ayman El-Baz, et al. MRBrainS Challenge: Online evaluation framework for brain image segmentation in 3T MRI scans. *Computational Intelligence and Neuroscience*, 2015.

[196] Mevislab. MeVisLab. http://www.mevislab.de/. [accessed April 5, 2016].

[197] Tom R Miller and Perry W Grigsby. Measurement of tumor volume by PET to evaluate prognosis in patients with advanced cervical cancer treated by radiation therapy. *International Journal of Radiation Oncology Biology Physics*, 53(2): 353–359, 2002.

[198] B Modayur, J Prothero, G Ojemann, K Maravilla, and J Brinkley. Visualization-based mapping of language function in the brain. *NeuroImage*, 6(4):245–258, 1997.

[199] Pim Moeskops, Max A Viergever, Manon JNL Benders, and Ivana Išgum. Evaluation of an automatic brain segmentation method developed for neonates on adult MR brain images. In *SPIE Medical Imaging*, pages 941315–941315. International Society for Optics and Photonics, 2015.

[200] Mahdi Mohamadkhanloo, Farzad Mehrabi, Abdolhamid Sohrabi, et al. Automatic determination of MS lesion subtypes based on fractal analysis in brain MR images. *Journal of Biomedical Science and Engineering*, 5(04):162, 2012.

[201] OJ Morris, M de J Lee, and AG Constantinides. Graph theory for image analysis: an approach based on the shortest spanning tree. *Communications, Radar and Signal Processing, IEE Proceedings F*, 133(2):146–152, 1986.

[202] MRI head. Magnetic Resonance Imaging (MRI) - Head. http://www.radiologyinfo.org/en/info.cfm?pg=headmr. [accessed April 5, 2016].

[203] MRI scanner. Magnetic Resonance Imaging. http://www.cyberphysics.co.uk/topics/medical/MRI.htm. [accessed April 1, 2016].

[204] Wayne Niblack. *An introduction to digital image processing*. Strandberg Publishing Company, 1985.

[205] Lars-Göran Nilsson and Hans J Markowitsch. *Cognitive neuroscience of memory*. Hogrefe & Huber Pub, 1999.

[206] American College of Radiology. Committee on drugs and contrast media. *ACR Manual on Contrast Media Version 10.1*, 9, 2013.

[207] Wonho Oh and W Brent Lindquist. Image thresholding by indicator kriging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):590–602, 1999.

[208] Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

[209] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[210] Dimitrios Pantazis, Anand Joshi, Jintao Jiang, David W Shattuck, Lynne E Bernstein, Hanna Damasio, and Richard M Leahy. Comparison of landmark-based and automatic methods for cortical surface registration. *NeuroImage*, 49 (3):2479–2493, 2010.

[211] Xenophon Papademetris, Andrea P Jackowski, Robert T Schultz, Lawrence H Staib, and James S Duncan. Integrated intensity and point-feature nonrigid registration. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004*, pages 763–770. 2004.

[212] Nikos Paragios. A variational approach for the segmentation of the left ventricle in cardiac image analysis. *International Journal of Computer Vision*, 50(3): 345–362, 2002.

[213] Jong Geun Park and Chulhee Lee. Skull stripping based on region growing for magnetic resonance brain images. *NeuroImage*, 47(4):1394–1407, 2009.

[214] Tamar Peli. An algorithm for recognition and localization of rotated and scaled objects. *USAF-NASA-sponsored research*, 69:483–485, 1981.

[215] Bo Peng, Lei Zhang, and David Zhang. A survey of graph theoretical approaches to image segmentation. *Pattern Recognition*, 46(3):1020–1038, 2013.

[216] Sérgio Pereira, Joana Festa, José António Mariz, Nuno Sousa, and Carlos A Silva. Automatic brain tissue segmentation of multi-sequence MR images using random decision forests. *Proceedings of the MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS'13)*, 2013.

[217] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[218] PET. Positron emission tomography. https://en.wikipedia.org/wiki/Positron_emission_tomography. [accessed January 26, 2016].

[219] Dzung L Pham, Chenyang Xu, and Jerry L Prince. Current methods in medical image segmentation 1. *Annual Review of Biomedical Engineering*, 2(1):315–337, 2000.

[220] DL Pope, Dennis L Parker, DE Gustafson, and PD Clayton. Dynamic search algorithms in left ventricular border recognition and analysis of coronary arteries. In *Computers in Cardiology*, pages 71–75, 1984.

[221] Gheorghe Postelnicu, Lilla Zöllei, and Bruce Fischl. Combined volumetric and surface registration. *IEEE Transactions on Medical Imaging*, 28(4):508–522, 2009.

[222] Marcel Prastawa, John H Gilmore, Weili Lin, and Guido Gerig. Automatic segmentation of MR images of the developing newborn brain. *Medical image analysis*, 9(5):457–466, 2005.

[223] William H Press. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press, 2007.

[224] C Pritchard, A Mayers, and D Baldwin. Changing patterns of neurological mortality in the 10 major developed countries–1979–2010. *Public Health*, 127 (4):357–368, 2013.

[225] Thierry Pun. Entropic thresholding, a new approach. *Computer Graphics and Image Processing*, 16(3):210–239, 1981.

[226] Jagath C Rajapakse and Frithjof Kruggel. Segmentation of MR images with intensity inhomogeneities. *Image and Vision Computing*, 16(3):165–180, 1998.

[227] Sanjay Ranade and Azriel Rosenfeld. Point pattern matching by relaxation. *Pattern Recognition*, 12(4):269–275, 1980.

[228] Richard A Redner and Homer F Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239, 1984.

[229] Kelly Rehm, Kirt Schaper, Jon Anderson, Roger Woods, Sarah Stoltzner, and David Rottenberg. Putting our heads together: a consensus approach to brain/non-brain segmentation in T1-weighted MR volumes. *NeuroImage*, 22 (3):1262–1270, 2004.

[230] Chantal Revol and Michel Jourlin. A new minimum variance region growing algorithm for image segmentation. *Pattern Recognition Letters*, 18(3):249–258, 1997.

[231] Chantal Revol-Muller, Francoise Peyrin, Yannick Carrillon, and Christophe Odet. Automated 3D region growing algorithm based on an assessment function. *Pattern Recognition Letters*, 23(1):137–150, 2002.

[232] TW Ridler and S Calvard. Picture thresholding using an iterative selection method. *IEEE Transactions on Systems, Man, and Cybernetics*, 8(8):630–632, 1978.

[233] Gustavo K Rohde, Akram Aldroubi, and Benoit M Dawant. The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Transactions on Medical Imaging*, 22(11):1470–1479, 2003.

[234] Torsten Rohlfing. Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. *IEEE Transactions on Medical Imaging*, 31(2):153–163, 2012.

[235] Torsten Rohlfing and Calvin R Maurer Jr. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Transactions on Information Technology in Biomedicine*, 7(1): 16–25, 2003.

[236] Karl Rohr, H Siegfried Stiehl, Rainer Sprengel, Thorsten M Buzug, Jürgen Weese, and MH Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on Medical Imaging*, 20(6):526–534, 2001.

[237] Serge ARB Rombouts, Frederik Barkhof, and Philip Scheltens. *Clinical applications of functional brain MRI*. Oxford University Press, 2007.

[238] Su Ruan, Cyril Jaggi, Jinghao Xue, Jalal Fadili, and Daniel Bloyet. Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE Transactions on Medical Imaging*, 19(12):1179–1187, 2000.

[239] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18 (8):712–721, 1999.

[240] Amit Ruf, Hayit Greenspan, and Jacob Goldberger. Tissue classification of noisy MR brain images using constrained GMM. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, volume 3750 of *Lecture Notes in Computer Science*, pages 790–797. Springer, 2005.

[241] Sachin D Ruikar and Dharmpal D Doye. Wavelet based image denoising technique. *International Journal of Advanced Computer Science and Applications*, 2(3):49–53, 2011.

[242] Detlef Ruprecht, Ralf Nagel, and Heinrich Müller. Spatial free-form deformation with scattered data interpolation methods. *Computers and Graphics*, 19(1): 63–71, 1995.

[243] Mert R Sabuncu, BT Thomas Yeo, Koen Van Leemput, Bruce Fischl, and Polina Golland. A generative model for image segmentation based on label fusion. *IEEE Transactions on Medical Imaging*, 29(10):1714–1729, 2010.

[244] Suresh A Sadananthan, Weili Zheng, Michael WL Chee, and Vitali Zagorodnov. Skull stripping using graph cuts. *NeuroImage*, 49(1):225–239, 2010.

[245] Prasanna K Sahoo, S Soltani, and Andrew KC Wong. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41(2):233–260, 1988.

[246] Mark J Sands and Abraham Levitin. Basics of magnetic resonance imaging. In *Seminars in Vascular Surgery*, volume 17, pages 66–82. Elsevier, 2004.

[247] Jose Santamaría, Oscar Cordón, and Sergio Damas. A comparative study of state-of-the-art evolutionary image registration methods for 3D modeling. *Computer Vision and Image Understanding*, 115(9):1340–1354, 2011.

[248] Michaël Sdika. Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote. *Medical Image Analysis*, 14(2):219–226, 2010.

[249] F Segonne, AM Dale, E Busa, M Glessner, D Salat, HK Hahn, and B Fischl. A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, 22(3): 1060–1075, 2004.

[250] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[251] David W Shattuck and Richard M Leahy. Brainsuite: an automated cortical surface identification tool. *Medical Image Analysis*, 6(2):129–142, 2002.

[252] David W Shattuck, Stephanie R Sandor-Leahy, Kirt A Schaper, David A Rottenberg, and Richard M Leahy. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13(5):856–876, 2001.

[253] David W Shattuck, Mubeena Mirza, Vitria Adisetiyo, Cornelius Hojatkashani, Georges Salamon, Katherine L Narr, Russell A Poldrack, Robert M Bilder, and Arthur W Toga. Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage*, 39(3):1064–1080, 2008.

[254] Dinggang Shen and Christos Davatzikos. HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, 2002.

[255] Dinggang Shen and Christos Davatzikos. Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration. *NeuroImage*, 18(1):28–41, 2003.

[256] Shan Shen, William Sandham, Malcolm Granat, and Annette Sterr. MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. *IEEE transactions on information technology in biomedicine*, 9 (3):459–467, 2005.

[257] Carolin J Sheridan, Tamara Matuz, Rossitza Draganova, Hari Eswaran, and Hubert Preissl. Fetal magnetoencephalography—achievements and challenges in the study of prenatal and early postnatal brain responses: a review. *Infant and Child Development*, 19(1):80–93, 2010.

[258] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[259] Kaleem Siddiqi, Bérubé Lauziere, Allen Tannenbaum, and Steven W Zucker. Area and length minimizing flows for shape segmentation. *IEEE Transactions on Image Processing*, 7(3):433–443, 1998.

[260] Adil Masood Siddiqui, Asif Masood, and Muhammad Saleem. A locally constrained radial basis function for registration and warping of images. *Pattern Recognition Letters*, 30(4):377–390, 2009.

[261] Sandro Sironi, Alessandro Buda, Maria Picchio, Patrizia Perego, Rossella Moreni, Antonio Pellegrino, Mario Colombo, Costantino Mangioni, Cristina Messa, and Ferruccio Fazio. Lymph node metastasis in patients with clinical early-stage cervical cancer: Detection with integrated FDG PET/CT 1. *Radiology*, 238(1):272–279, 2006.

[262] John G Sled, Alex P Zijdenbos, and Alan C Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17(1):87–97, 1998.

[263] Stephen M Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, 2002.

[264] Stephen M Smith, Yongyue Zhang, Mark Jenkinson, Jacqueline Chen, PM Matthews, Antonio Federico, and Nicola De Stefano. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage*, 17(1):479–489, 2002.

[265] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, Rami K Niazy, James Saunders, John Vickers, Yongyue Zhang, Nicola De Stefano, J Michael Brady, and Paul M Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23:S208–S219, 2004.

[266] Marine Soret, Stephen L Bacharach, and Irene Buvat. Partial-volume effect in PET tumor imaging. *Journal of Nuclear Medicine*, 48(6):932–945, 2007.

[267] Aristeidis Sotiras, Nikos Komodakis, Ben Glocker, Jean-François Deux, and Nikos Paragios. Graphical models and deformable diffeomorphic population registration using global and local metrics. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pages 672–679. Springer, 2009.

[268] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32 (7):1153–1190, 2013.

[269] SPECT. Medscape: SPECT Brain Imaging. http://emedicine.medscape.com/article/2064780-overview#a5, . [accessed April 5, 2016].

[270] SPECT. Nuclear medicine of Naples, L.L.C., division of nuclear neurology, what is a brain SPET scan?, .

[271] George Stockman, Steven Kopstein, and Sanford Benett. Matching images to models for registration and object detection via clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (3):229–241, 1982.

[272] Marijn F Stollenga, Wonmin Byeon, Marcus Liwicki, and Juergen Schmidhuber. Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation. In *Advances in Neural Information Processing Systems*, pages 2980–2988, 2015.

[273] Heung-Il Suk and Dinggang Shen. Deep learning-based feature representation for AD/MCI classification. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 583–590. Springer, 2013.

[274] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1891–1898, 2014.

[275] Johan AK Suykens. *Advances in learning theory: methods, models, and applications*, volume 190. IOS Press, 2003.

[276] László Szilagyi, Zoltán Benyo, Sándor M Szilágyi, and HS Adam. MR brain image segmentation using an enhanced fuzzy c-means algorithm. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2003.*, volume 1, pages 724–726. IEEE, 2003.

[277] Tommy WH Tang and Albert CS Chung. Non-rigid image registration using graph-cuts. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pages 916–924. Springer, 2007.

[278] Xiaodong Tao and Ming-Ching Chang. A skull stripping method using deformable surface and tissue classification. In *SPIE Medical Imaging*, pages 76233L–76233L. International Society for Optics and Photonics, 2010.

[279] Igor V Tetko, David J Livingstone, and Alexander I Luik. Neural network studies. 1. comparison of overfitting and overtraining. *Journal of Chemical Information and Computer Sciences*, 35(5):826–833, 1995.

[280] Philippe Thévenaz, Thierry Blu, and Michael Unser. Image interpolation and resampling. *Handbook of Medical Imaging, Processing and Analysis*, pages 393–420, 2000.

[281] Jean-Philippe Thirion. Fast non-rigid matching of 3D medical images. *[Research Report] RR-2547*, page 37, 1995.

[282] Jean-Philippe Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

[283] Dan Tian and Linan Fan. A brain MR images segmentation method based on SOM neural network. In *2007 1st International Conference on Bioinformatics and Biomedical Engineering*, pages 686–689. IEEE, 2007.

[284] Arthur W Toga and Paul M Thompson. The role of image registration in brain mapping. *Image and Vision Computing*, 19(1):3–24, 2001.

[285] Jussi Tohka, Evgeny Krestyannikov, Ivo D Dinov, Allan MacKenzie Graham, David W Shattuck, Ulla Ruotsalainen, and Arthur W Toga. Genetic algorithms for finite mixture model based voxel classification in neuroimaging. *IEEE transactions on medical imaging*, 26(5):696–711, 2007.

[286] Jussi Tohka, Ivo D Dinov, David W Shattuck, and Arthur W Toga. Brain MRI tissue classification based on local markov random fields. *Magnetic Resonance Imaging*, 28(4):557–573, 2010.

[287] Alain Trouvé. Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision*, 28(3):213–221, 1998.

[288] Srinivas C Turaga, Joseph F Murray, Viren Jain, Fabian Roth, Moritz Helmstaedter, Kevin Briggman, Winfried Denk, and H Sebastian Seung. Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural computation*, 22(2):511–538, 2010.

[289] US. Radiology Assistant: Neonatal Brain US. http://www.radiologyassistant.nl/en/p440c93be7456f/neonatal-brain-us.html. [accessed Febuary 19, 2016].

[290] Sergi Valverde, Arnau Oliver, Mariano Cabezas, Eloy Roura, and Xavier Lladó. Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations. *Journal of Magnetic Resonance Imaging*, 41(1):93–101, 2015.

[291] André JW van der Kouwe, Thomas Benner, David H Salat, and Bruce Fischl. Brain morphometry with multiecho mprage. *NeuroImage*, 40(2):559–569, 2008.

[292] Fedde van der Lijn, Tom den Heijer, Monique MB Breteler, and Wiro J Niessen. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage*, 43(4):708–720, 2008.

[293] Fedde Van der Lijn, Marleen De Bruijne, Stefan Klein, Tom Den Heijer, Yoo Y Hoogendam, Aad Van der Lugt, Monique Breteler, and Wiro J Niessen. Automated brain structure segmentation based on atlas registration and appearance models. *IEEE Transactions on Medical Imaging*, 31(2):276–286, 2012.

[294] Annegreet van Opbroek, Fedde van der Lijn, and Marleen de Bruijne. Automated brain-tissue segmentation by multi-feature SVM classification. *MICCAI Grand Challenge on MR Brain Image Segmentation workshop*, 2013.

[295] Annegreet van Opbroek, Fedde van der Lijn, and Marleen de Bruijne. Automated brain-tissue segmentation by multi-feature SVM classification. 2013.

[296] Baba C Vemuri, Shuangying Huang, Sartaj Sahni, Christiana M Leonard, Cecile Mohr, Robin Gilmore, and Jeffrey Fitzsimmons. An efficient motion estimator with application to medical image registration. *Medical Image Analysis*, 2(1): 79–98, 1998.

[297] Baba C Vemuri, J Ye, Y Chen, and Christiana Morison Leonard. Image registration via level-set motion: Applications to atlas-based segmentation. *Medical Image Analysis*, 7(1):1–20, 2003.

[298] Prashanthi Vemuri, Jeffrey L Gunter, Matthew L Senjem, Jennifer L Whitwell, Kejal Kantarci, David S Knopman, Bradley F Boeve, Ronald C Petersen, and Clifford R Jack. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *NeuroImage*, 39(3):1186–1197, 2008.

[299] Uroš Vovk, Franjo Pernuš, and Boštjan Likar. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Transactions on Medical Imaging*, 26(3):405–421, 2007.

[300] H Vrooman, Fedde van der Lijn, and W Niessen. Auto-kNN: brain tissue segmentation using automatically trained k-nearest-neighbor classification. In *Proceedings of the MICCAI Workshops—The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS'13)*, 2013.

[301] Henri A Vrooman, Chris A Cocosco, Rik Stokking, M Arfan Ikram, Meike W Vernooij, Monique M Breteler, and Wiro J Niessen. kNN-based multi-spectral MRI brain tissue classification: manual training versus automated atlas-based training. In *SPIE Medical Imaging*, pages 61443L–61443L. International Society for Optics and Photonics, 2006.

[302] Henri A Vrooman, Chris A Cocosco, Fedde van der Lijn, Rik Stokking, M Arfan Ikram, Meike W Vernooij, Monique MB Breteler, and Wiro J Niessen. Multi-spectral brain tissue segmentation using automatically trained k-nearest-neighbor classification. *NeuroImage*, 37(1):71–81, 2007.

[303] C Wachinger, M Toews, G Langs, W Wells, and P Golland. Keypoint transfer segmentation. In *Information Processing in Medical Imaging*, volume 9123 of *Lecture Notes in Computer Science*, pages 233–245. Springer, 2015.

[304] He Wang, Lei Dong, Jennifer O'Daniel, Radhe Mohan, Adam S Garden, K Kian Ang, Deborah A Kuban, Mark Bonnen, Joe Y Chang, and Rex Cheung. Validation of an accelerated'demons' algorithm for deformable image registration in radiation therapy. *Physics in Medicine and Biology*, 50(12):2887–2905, 2005.

[305] Hongzhi Wang, Jung Wook Suh, John Pluta, Murat Altinay, and Paul Yushkevich. Optimal weights for multi-atlas label fusion. In *Information Processing*

in Medical Imaging, volume 6801 of *Lecture Notes in Computer Science*, pages 73–84. Springer, 2011.

[306] Li Wang, Feng Shi, Yaozong Gao, Gang Li, John H Gilmore, Weili Lin, and Dinggang Shen. Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain mr image segmentation. *NeuroImage*, 89:152–164, 2014.

[307] Li Wang, Yaozong Gao, Feng Shi, Gang Li, John H Gilmore, Weili Lin, and Dinggang Shen. LINKS: Learning-based multi-source integration framework for segmentation of infant brain images. *NeuroImage*, 108:160–172, 2015.

[308] Liping Wang, Ziming Zeng, and Reyer Zwiggelaar. An improved BET method for brain segmentation. In *22nd International Conference on Pattern Recognition (ICPR), 2014*, pages 3221–3226. IEEE, 2014.

[309] Shyuan Wang and Robert M Haralick. Automatic multithreshold selection. *Computer Vision, Graphics, and Image Processing*, 25(1):46–67, 1984.

[310] Song Wang and Jeffrey Mark Siskind. Image segmentation with minimum mean cut. In *Eighth IEEE International Conference on Computer Vision, 2001. Proceedings.*, volume 1, pages 517–524. IEEE, 2001.

[311] Song Wang and Jeffrey Mark Siskind. Image segmentation with ratio cut. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):675–690, 2003.

[312] Xun Wang, Lei He, and William Wee. Deformable contour method: a constrained optimization approach. *International Journal of Computer Vision*, 59 (1):87–108, 2004.

[313] Simon K Warfield, Kelly H Zou, and William M Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004.

[314] Neil I Weisenfeld, Andrea UJ Mewes, and Simon K Warfield. Highly accurate segmentation of brain tissue and subcortical gray matter from newborn MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 199–206. Springer, 2006.

[315] Holger Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4(1):389–396, 1995.

[316] Roger P Woods, Scott T Grafton, Colin J Holmes, Simon R Cherry, and John C Mazziotta. Automated image registration: I. general methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, 22(1): 139–152, 1998.

[317] Roger P Woods, Scott T Grafton, John DG Watson, Nancy L Sicotte, and John C Mazziotta. Automated image registration: II. intersubject validation of linear and nonlinear models. *Journal of Computer Assisted Tomography*, 22(1): 153–165, 1998.

[318] Angela Y Wu, Tsai-Hong Hong, and Azriel Rosenfeld. Threshold selection using quadtrees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):90–94, 1982.

[319] Guorong Wu, Minjeong Kim, Qian Wang, and Dinggang Shen. S-HAMMER: Hierarchical attribute-guided, symmetric diffeomorphic registration for MR brain images. *Human Brain Mapping*, 35(3):1044–1060, 2014.

[320] Guorong Wu, Qian Wang, Daoqiang Zhang, Feiping Nie, Heng Huang, and Dinggang Shen. A generative probability model of joint label fusion for multi-atlas based brain segmentation. *Medical Image Analysis*, 18(6):881–890, 2014.

[321] Ying Wu, Simon K Warfield, I Leng Tan, William M Wells, Dominik S Meier, Ronald A van Schijndel, Frederik Barkhof, and Charles RG Guttmann. Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. *NeuroImage*, 32(3):1205–1215, 2006.

[322] Zhenyv Wu and Richard Leahy. Tissue classification in MR images using hierarchical segmentation. In *Nuclear Science Symposium, 1990. Conference record: Including Sessions on Nuclear Power Systems and Medical Imaging Conference*, pages 1410–1414. IEEE, 1990.

[323] Zhao Yi, Antonio Criminisi, Jamie Shotton, and Andrew Blake. Discriminative, semantic segmentation of brain tissue in MR images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, volume 5762 of *Lecture Notes in Computer Science*, pages 558–565. Springer, 2009.

[324] Mehmet Yigitsoy, Christian Wachinger, and Nassir Navab. Temporal groupwise registration for motion modeling. In *Information Processing in Medical Imaging*, pages 648–659. Springer, 2011.

[325] Hideteru Yoshida. Effect of spatial resolution on SPECT quantffication values. *The Journal of Nuclear Medicine*, 30:508–514, 1989.

[326] Sahar Yousefi, Reza Azmi, and Morteza Zahedi. Brain tissue segmentation in MR images based on a hybrid of MRF and social algorithms. *Medical Image Analysis*, 16(4):840–848, 2012.

[327] Evangelia I Zacharaki, Stathis Kanterakis, R Nick Bryan, and Christos Davatzikos. Measuring brain lesion progression with a supervised tissue classification system. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, volume 5241 of *Lecture Notes in Computer Science*, pages 620–627. Springer, 2008.

[328] Charles T Zahn. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers*, 100(1):68–86, 1971.

[329] Dao-Qiang Zhang and Song-Can Chen. A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. *Artificial Intelligence in Medicine*, 32(1):37–50, 2004.

[330] Hui Zhang, Jason E Fritts, and Sally A Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260–280, 2008.

[331] Shanrong Zhang, Matthew Merritt, Donald E Woessner, Robert E Lenkinski, and A Dean Sherry. Paracest agents: modulating MRI contrast via water proton exchange. *Accounts of Chemical Research*, 36(10):783–790, 2003.

[332] Yongyue Zhang, Michael Brady, and Stephen Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.

[333] Yu Jin Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.

[334] Yudong Zhang, Zhengchao Dong, Lenan Wu, and Shuihua Wang. A hybrid method for MRI brain image classification. *Expert Systems with Applications*, 38(8):10049–10053, 2011.

[335] Qinfen Zheng and Rama Chellappa. A computational vision approach to image registration. *IEEE Transactions on Image Processing*, 2(3):311–326, 1993.

[336] Audrey H Zhuang, Daniel J Valentino, and Arthur W Toga. Skull-stripping magnetic resonance brain images using a model-based level set. *NeuroImage*, 32(1):79–92, 2006.

[337] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

[338] Torre D Zuk, M Stella Atkins, and Kellogg S Booth. Approaches to registration using 3D surfaces. In *Medical Imaging 1994*, pages 176–187. International Society for Optics and Photonics, 1994.